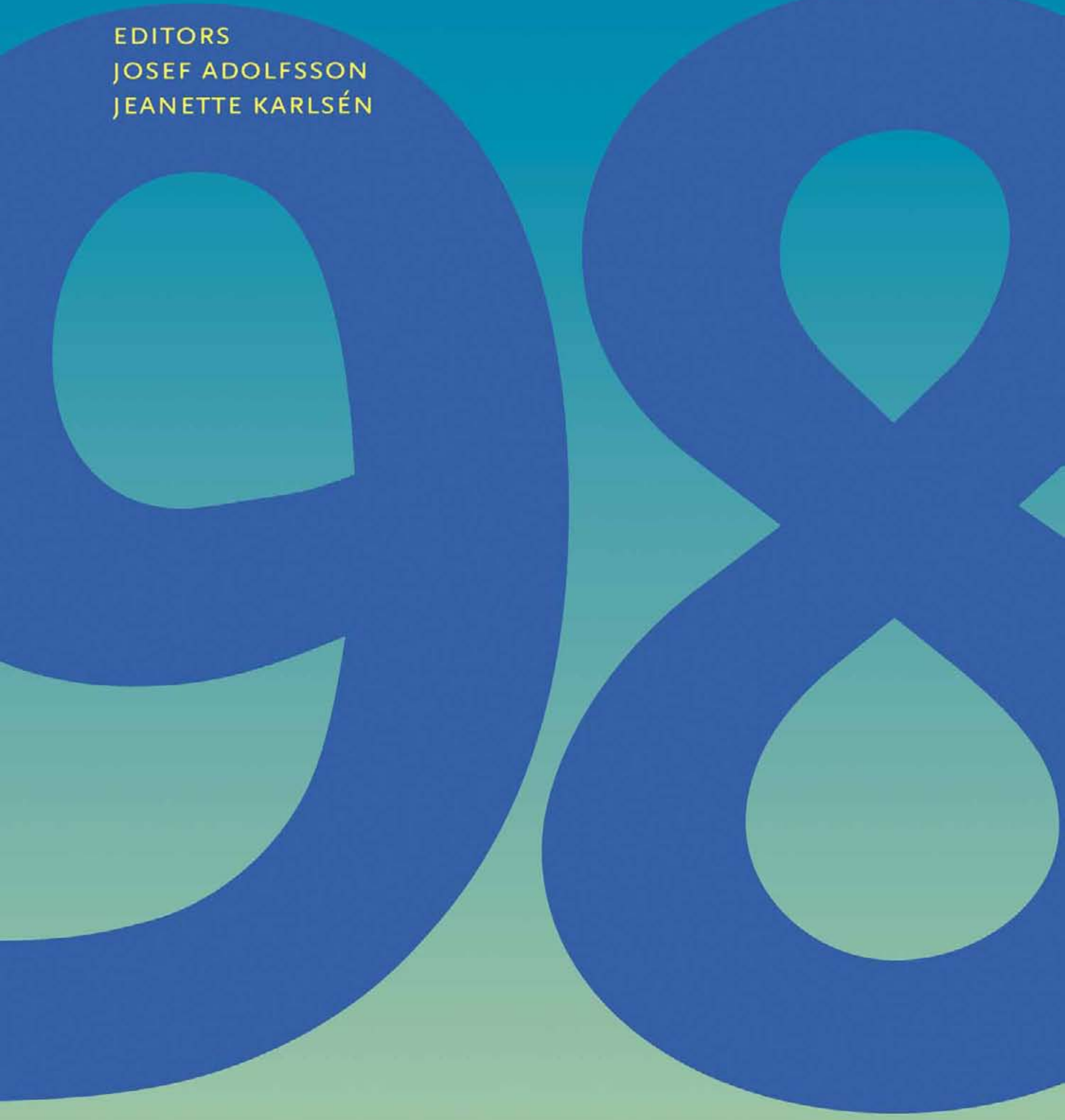
MECHATRONICS '98

EDITORS
JOSEF ADOLFSSON
JEANETTE KARLSÉN



PERGAMON

VISIT…

LANZAROTE Caliente.COM

MECHATRONICS '98

Proceedings of the 6th UK Mechatronics Forum International Conference

Elsevier Science Internet Homepage - http://www.elsevier.nl
Contains full catalogue information on all books, journals and electronic products

Elsevier Titles of Related Interest

BONIS ET AL.
International Progress in Precision Engineering

KWAY & TANAKA
Computational Engineering

SODHI
Advances in Manufacturing Systems: Design, Modeling and Analysis

STOLARSKI
Analytical Design Techniques

USUI
Advancement of Intelligent Production

Related Journals

Computer Methods in Applied Mechanics and Engineering
Computer-Aided Design
Computers and Electrical Engineering
International Journal of Engineering Science
International Journal of Mechanical Sciences
Mechanism and Machine Theory
Mechatronics
Precision Engineering
Robotics and Autonomous Systems
Robotics and Computer-Integrated Manufacturing
Sensors and Actuators A: Physical

MECHATRONICS '98

Proceedings of the 6th UK Mechatronics Forum International Conference,
Skövde, Sweden, 9 - 11 September 1998

Editors
Josef Adolfsson and Jeanette Karlsén
University of Skövde, Skövde, Sweden

Organised by
UK Mechatronics Forum
Centre for Intelligent Automation, University of Skövde
De Montford University

Sponsored by

ABB Flexible Automation
CEJN
FESTO
Volvo
IEE, Manufacturing Division
IMechE, Machine Systems Computing and Control
Prosolvia Systems AB
Skövde Kommun

Pergamon
Amsterdam – Lausanne – New York – Oxford – Shannon – Singapore – Tokyo

ELSEVIER SCIENCE Ltd, The Boulevard, Langford Lane,
Kidlington, Oxford, OX5 1GB, UK

Library of Congress Cataloging-in-Publication Data
UK Mechatronics Forum International Conference (6th : 1998 : Skövde, Sweden)
Mechatronics '98 : proceedings of the 6th UK Mechatronics Forum International Conference, Skövde, Sweden, 9–11 September 1998 / editors, Josef Adolfsson and Jeanette Karlsén : organised by UK Mechatronics Forum, Centre for Intelligent Automation, University of Skövde, De Montfort University : sponsored by ABB Fle
Automation ... [et al.]. — 1st ed.
p. cm.
ISBN 0-08-043339-1 (hardcover)
1. Mechatronics—Congresses.
I. Adolfsson, Josef. II. Karlsén, Jeanette. III. Högskolan i Skövde. Centre for Intelli
Automation. IV. De Montfort University. V. Title.
TJ163.12.U35 1998
621—dc21 98-29168
CIP

British Library Cataloguing in Publication data
A catalogue record for this title is available from the British Library

Copyright © 1998 Elsevier Science Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronics, electrostatic, magnetic tape, mechanical photocopying, recording or otherwise, without prior permission in writing from the copyright holder.

First edition 1998

ISBN: 0-08-043339-1

♾The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).

Printed in The Netherlands.

Foreword

A1. Control and its Applications

Control of flexible assembly systems with a fault recovery capability
W H R Yeung, P R Moore

Development of an open embedded machine controller for the automotive industry
R Harrison, L Lee

A Petri net-based inference network for design automation under nondeterminism applied to mechatronic systems
Z Erden, A M Erkmen, A Erden

Towards paradigm shift in machine design and control
J Pu, P R Moore

Modeling of failure treatment in manufacturing systems through Petri nets
P E Miyagi, L A M Riascos, N Maruyama, F G Cozman

A2. Control and its Applications

Synthesis and analysis of feedback linearization for pneumatic actuator control systems
F Xiang, J Wikander

On line autotuning for fuzzy logic controlled pneumatic positioners
G Mattiazzo, S Mauro, M Velardocchia, T Raparelli

An experimental workbench for innovative pneumatic units cooperant under distributed control
E Ravina, R Fattori

Development of a new electronic-pneumatic control for pneumatic systems
G Belforte, G Eula, C Ferraresi, V Viktorov

Smart components-based servo pneumatic actuation systems
J Pu, P R Moore, C B Wong, J H Wang

A3. Control and its Applications

A modelling framework for design and analysis of distributed real-time control implementations 67
O Redell, M Törngren

Robust motion control for planar laser cutting machine 73
A Hace, K Jezernik, B Curk, M Terbuc

Controlling a shape memory alloy drived robot gripper by fuzzy logic 79
I Mihálcz, P Baranyi, Z Balogh, P Korondi, L Valenta, A Halmai

Simulation of an automatic vehicle transmission as a mechatronic system 85
A Haj-Fraj, F Pfeiffer

Resonance and quasi-resonance drives for start-stop regime 91
T Akinfiev, M Armada

Hardware oriented modeling and system identification: breaking the barrier between design and analysis 97
M Donnelly, J Langenwalter

A4. Control and its Applications

A high-precision positioning servo-controller using non-sinusoidal two-phase type PLL 103
L Wang, T Emura

Modeling and adaptive control of permanent magnet synchronous motors using multilayer neural networks 109
A Albostan, M Gökbulut

Robust position control for a DC servomechanism subject to friction and compliance 117
Y Li, B Eriksson, J Wikander

Mechanical object driven by the linear DC motor. Modeling, identification and control concepts 123
J Gustowski

Determination of the transfer function for the mechanical objects along transitional response 129
K Sarkauskas, R Dapkus, M Rackaitis

A5. Mechatronics in Production

Smooth trajectory planning for planar laser cutting system 135
M Rodic, A Hace, M Terbuc, K Jezernik

Adaptive control of mechanical processes: brakeforming of metal sheet as an example 141
L J de Vin, U P Singh

A piezoelectrically driven wire feeding system for high performance wedge-wedge-bonding machines 147
A Henke, M A Kümmel, J Wallaschek

Behaviour-based assembly of a family of electric torches 153
G C Pettinaro

A6. Mechatronics in Production

Electronic reconfiguration in flexible winding type machines 159
N Jakeman, W A Bullough, P H Mellor

Mechatronic design of a leather spray system 167
S Y T Lang, K P Rao, L T Ching

A mechatronic system for the improvement of surface form in planed and moulded timber components 173
N Brown, R M Parkin

A7. Control and its Applications

Application of fuzzy PI control to improve the positioning accuracy of a rotary-linear motor driven by two-dimensional ultrasonic actuators 181
M Kinouchi, I Hayashi, N Iwatsuki, K Morikawa, J Shibata, K Suzuki

Mechatronic systems with uncertain physical parameters 187
E Coelingh, T J A de Vries, J Holterman, J van Amerongen

The design and implementation of a surface oriented controller 193
A J Winsby, S E Burge, I A Grout

Mechatronics for moving machinery - a precision, multi-axis, high speed and acceleration winch system 199
N Darracott, M Honeywill

Mechatronics in medical engineering: advanced control of a ventilation device 205
I Jenayeh, F Simon, H Rake

A8. Control and its Applications

Modeling of a distributed, non-steady state, thermal system for the purpose of control 211
M M Chen, K C Craig, G A Domoto

Experiment to evaluate the feasibility of the smart disc concept 217
J Holterman, T J A de Vries, M P Koster

Modelling and motion control of an autonomous underwater robot using a fuzzy and fuzzy neural approach 223
I S Akkizidis, G N Roberts

A simplified fuzzy logic control scheme 229
C M Lim, Y Z Ouyang

Control strategies of injection during operation of S.I. engines 235
J J G Martins

B1. Robotics

A flexible end-effector for robotized assembly 241
J A N Loureiro , M P F S Gomes

Mechatronical design of underwater sensor/actuator robots for cooperative task execution 249
P Appelqvist, M Vainio, A Halme

Neural network adaptive controller for DD robot 255
R Safaric, K Jezernik, M Kupljen, R Parkin

Adaptive tracking error controller for rigid-link flexible-joint robot manipulators 261
C J Tsaprounis, N Aspragathos

Obstacle avoidance control based on an harmonic potential field 267
S A Krasnova

B2. Robotics

Smooth robust control for robot manipulators 275
B W Bekit, J F Whidborne, L D Seneviratne

Robot arm positional deflection control with a laser light sensor 281
M G Puopolo, S B Niku

The control of elastic multi-link manipulator based on the dynamic compensation method 287
V A Utkin

Drive dimension and path preparation for parallel manipulators 293
T L Tran, G Kehl

Passivity based adaptive controller for a two degrees of freedom robot using a composite parameters adaptation law 299
S Dumbrava

B3. Mobile Robots and AGVs

New workhorse for lumberjacks 305
P Vähä, K Känsälä, J Kerva, P Saavalainen

Intelligent semi-autonomous vehicles in materials handling 311
P R Moore, J Pu, J-O Lundgren, S Ujvari

Remote radiation mapping and preliminary intervention using collaborating (European and Russian) mobile robots 317
L Piotrowski, E Loane, M Halbach, N Sidorkin

Development of mobile robots for mines detection 323
E Colon, Y Baudoin, P Alexandre

B4. Mobile Robots and AGVs

Architecture for embedded systems in vehicles and mobile machinery 329
B T Henoch

Towards temporally embedded systems: dynamic short-term memory for mobile robots 337
R M Rylatt, C A Czarnecki

Automatic guided vehicle with independent rear wheel DC motor drivers and front wheel steering 343
V González, J Guadarrama, M Lopez

An intuitive control and sensory system for the enhancement of remotely controlled vehicle operation 349
P Webb, J I Robson

B5. Walking Robots

An intelligent mechatronic architecture for hexapod control 355
M J Randall, A G Pipe, A F T Winfield

Behavior-based control of a four legged walking robot 361
L Pettersson, K Jansson, H Rehbinder, J Wikander

Biologically inspired design and construction of a compliant robot leg for dynamic walking 367
F Hardarson, K Benjelloun, T Wadden, J Wikander

Adaptive neural control of hexapod leg trajectories 373
M J Randall, AG Pipe, A F T Winfield

Linear viscoelastic actuator-based control system of a bipedal walking robot 379
V Berbyuk, B Peterson, N Nishchenko

B6. Mobile Robots and AGVs

Simulation and emulation of sensor systems for intelligent vehicles 385
S Ujvari, P Eriksson, P R Moore, J Pu

Modified control system for mobile robots 391
N A Lakota, V S Lapshov

A dynamic agent architecture and token negotiations 397
A Wallace

Automatic navigation in a complex diffusion field environment 403
E E Kadar, G S Virk

B7. Mobile Robots and AGVs

Artificial evolution of street-crossing robots 409
M Wahde

Neuroprocessor control system of intelligent mobile robots 415
I Kaliaev, I Rubtsov, V Chelyshev

Trajectory control and time-varying stabilization of nonholonomic wheeled mobile robot: a hybrid strategy 421
A C Victorino, P R G Kurka, E G O Nobrega

Visual guidance of mobile robots using a neural network 427
A J Baerveldt, A Björnberg, M Gisbert

C1. Novel Sensors and Actuators

Control of piezo actuated mechanisms 433
J van Dijk

Resonantly driven piezoelectric micropump 441
J-H Park, K Yoshida, S Yokota

Piezoelectric final-control elements in mechatronic devices 447
A Boshliakov, A Dubin, I Rubtsov, A Shtcherbin

An SMA-driven micropump using electro-conjugate fluid jet cooling 453
S Yokota, K Yoshida, T Kawaue, Y Otsubo, K Edamura

Information transmitting properties of mechatronic systems 459
Z Kaposvári, J Rostás, T Szabó

C2. Novel Sensors and Actuators

New polarized electromagnetic actuators as integrated mechatronic components – design and application 465
E Kallenbach, H Kube, K Feindt, V Zöppig, R Hermann, F Beyer

The recent advantages of the axial flux DC micromotors 473
A Halmai

Fast actuation of a hydraulic switching valve by parametrically excited structures 477
M Garstenauer , R Scheidl

The piston-in-cylinder electromagnetic ram 483
P Denne

Smart ultrasonic sensor for semi-autonomous robots 489
R Kazys

C3. Design

Mechatronic development environment for precision engineering 495
H Grabowski, S Rude, Z Pocsai, M Huang

Mechatronic system design methodology – initial principles based on case studies 501
M Valásek

4D Design: adding value to consumer products using mechatronics and multimedia technologies 507
A Robertson

On synergistic aspects in integrated product development of mechatronic systems 513
F Kaljas, V Reedik

Design, fabrication, and testing of miniature polymer gripper systems 517
R J Chang, H S Wang, Y L Wang

C4. Design

Design methodology of application specific integrated circuits for mechatronic systems 523
F- M Renner, K-J Hoffmann, R Markert, M Glesner

Mechatronic design of digital electrovalves modulated in PWM 529
M Sorli, G Figliolini

A computer aided conceptual design method for mechatronic systems in process oriented heavy machinery 535
St Dierneder, R Scheidl, K Mörwald, J Guttenbrunner

A computer based tool to support electronics design in a mechatronic environment 541
R M Walters, D A Bradley, A P Dorey

A comparison between modelling approaches for custom electronic circuits within a control systems application 547
I A Grout, J Santana-Corte, A Winsby, S Burge

C5. Novel Sensors and Actuators

Wireless remote sensing and identification by using passive SAW devices 553
W Buff, M Goroll

Wireless identification and sensing using surface acoustic wave devices 559
A Springer, R Weigel, A Pohl, F Seifert

A Ka-band doppler-sensor 565
C G Diskus, A Stelzer, K Lübke, A L Springer, H W Thim

Measuring mechanical parameters utilizing wireless passive sensors 571
A Pohl, R Steindl, L Reindl, F Seifert

A new two-axis attitude sensor based on free pendulum inclinometer 577
N Getschko

C6. Vision in Real-Time Control

A vision assisted mechatronic system for arc welding 583
L G Trabasso, C E F Silva

The automated sorting of a bin of randomly placed documents of variable type and condition 589
C Jones, G Ross, J R Hewit

A low-cost colour vision-system for robot design competitions 595
A J Baerveldt, B Åstrand

Exploring spherical images properties from off-the-shelf cameras 601
I Fonseca, J Dias

Combining a fuzzy classifier with classifiers based on statistic moments 607
T G Amaral, M M Crisóstomo, A T de Almeida

Surface profile based on sensor fusion 613
C Santos, J Fonseca, P Garrido, C Couto

C7. Mechatronics In Non-Industrial Applications

Features of the lateral active pneumatic suspensions in the ETR 470 high speed train 621
M Sorli, W Franco, S Mauro, G Quaglia, R Giuzio, G Vernillo

Mechatronics in automating shunting processes in marshalling yards 627
I Jenayeh, Ch Müller, N Kositza, M Enning, H Rake

The development of a hardware-in-the-loop simulation of a pump storage hydro power station 633
S Mansoor, D Bradley, C Aris, G Jones

Mechanical earthworm to facilitate the access and repair to buried infrastructure 639
W L Dudeney, M R Jackson

D1. Mechatronics Education

Training of experts in mechatronics in Moscow Bauman State Technical University 645
N A Lakota, E D Gorbatchevich, V S Medvedev

Mechatronics as an engineering science 651
J Wikander, M Törngren

The Pacificar Project as an integral mechatronics course component: 657
A concept for linking educational and research activities in the field of mechatronics
C Brom

Mechatronics education at ETH Zurich based on 'hands on experience' 667
R Y Siegwart, R Büchi, P Bühler

Mechatronics education in unified curriculum 673
C-W Jo, S-T Kim, C-R Joe

D2. Mechatronics Education / Mechatronics in the Textile Industries

TESLA - a modular system for laboratory automation 679
M Robrecht

Mechatronics design of an inverted pendulum system for engineering education 685
C S Tüfekçi, J A de Marchi, K C Craig, S Özçelik

Enhancing a mechanical engineering curriculum with a mechatronics theme 693
D G Alciatore, M B Histand

Fuzzy control applications development system 697
J Alvarez , A Manzanedo

Tribological measurement of the state of surface fabrics by a contact and non-contact methods 703
M-A Bueno, M Renner, B Durand

A novel mechatronic approach in machine vision for lace inspection 709
H R Yazdi, T G King

D3. Mechatronics in the Textile Industries

Digital drive system for stitch length control in circular knitting 715
T Dias, C-F Tang, G. Lanarolle

Mechatronic principles for aesthetic measurement of textile materials 721
G K Stylios

A sewing rig with automatic feature extraction 727
H Carvalho, F N Ferreira, J L Monteiro, A M Rocha

Study of the feeding system of an overlock sewing machine 733
L F da Silva, M Lima, C Couto, F N Ferreira, D Andrade, L P Ferreira

Determining the effectiveness of computer vision when measuring yarn interlace at high speeds 739
M P Millman, M Acar, M R Jackson

In-process fabric defect detection and identification 745
J L Dorrity, G J Vachtsevanos

D4. Calibration, Measurement and Inspection

Estimation of workpiece quality by using state observer 751
M Shiraishi, S Aoshima

On the position and velocity measurements of the high speed piston of a free piston engine 757
E Lehto, T Virvalo, J Inberg

Inference mechanisms applied to on-line electro-pneumatic fault detection and diagnosis 763
D S Evans

An unsupervised artificial neural network approach to adaptive noise cancellation applied to on-line tool condition monitoring 769
M Girolami, J Findlay

Autonomous system for oil pipelines inspection 775
J Okamoto Jr, J C Adamowski, M S G Tsuzuki, F Buiochi, C S Camerini

D5. Virtual Reality / Teleoperation

A novel 3D input device, with 6 DoF, prototype based mechanically on a gyro and electronically via an FPGA 781
P E Jones

The application of a systems methodology to the design and specification of an intelligent telecare system 787
G Williams, D A Bradley, K Doughty

Redundant manipulators as a tool for disabled people in an unstructured environment 793
M Olsson, P Hedenborn, U Lorentzon, G Bolmsjö

A helmet mounted display system with active gaze control for visual telepresence 799
J P Brooker, P M Sharkey

Integrated telepresence, virtual reality, and mobile communications - A project report 805
G Mair, N Clark, R Fryer, R Hardiman, D McGregor, A Retik, N Retik, K Revie

D6. Teleoperation

Development of an electropneumatic device for remotely sensed manipulation 811
C Ferraresi, A M Bertetto, H R Z Niasar

A mechatronic approach to auditory localisation for transparent telepresence 817
C S Harrison, G M Mair

A novel teleoperated robot control system 823
A R Graves, C A Czarnecki

Control of the robot system via internet 829
R Safaric, D W Calkin, R M Parkin, C A Czarnecki

Man-machine interface for remote programming and control of NC machine tools at the task level: An example 835
F Moreira, G Putnik, R Sousa

Control methods for force reflective master slave systems 843
H Flemmer, B Eriksson, J Wikander

D7. System Modeling and Simulation

On modelling mechatronics systems 849
B A Hussein

Visual-MOOMo - a graphical object-oriented modelling environment for the design of mechatronic systems 855
M Wolf

A development strategy for mechatronic systems based on functional and geometrical modelling techniques 861
M Kümmel, A Henke, J Wallaschek

3-D graphical simulation for agile modular machine systems 867
J Adolfsson, P Olofsgård, P R Moore, J Pu

Object-oriented modeling and simulation of mechatronic systems with 20-sim 3.0 873
P B T Weustink, T J A de Vries, P C Breedveld

D8. System Modeling and Simulation

Integrated computer aided programming (CAP) system for robots in reverse engineering applications 879
P Leonov

Sensor controlled robotic arc welding integrated with simulation software 885
P Hedenborn, M Olsson, G Bolmsjö

Simulation of continuous process variations in production system design 891
J Oscarsson, P R Moore

Interfacing autolev to matlab and simulink 897
M Abderrahim, A R Whittaker

User-modelling to improve the usability of systems for computer-aided process planning 903
A R Nordqvist, P T Eriksson, P R Moore

Author index 911

FOREWORD FROM THE CO-CHAIRS

Mechatronics, although still a relatively new term when compared with many of the traditional engineering disciplines, now appears firmly established. Individuals, industries and Universities around the world are now using the term freely, in the understanding that it has an international currency. The Mechatronics Forum is proud to have played its part in establishing and promoting the merits of the integrated, multi-disciplinary, *mechatronic* approach in the research, development and design of new products and systems, partly through its sponsorship and organisation of its biennial conferences. The Mechatronics Forum has now been involved in promoting the discipline for over twelve years and is currently planning it seventh conference for the year 2000.

This years conference **Mechatronics '98** is notable in that, it runs in parallel with the ECDVRAT '98 Conference in one venue, with the consequence that it will gather a large international grouping of practitioners and experts to share ideas and discuss new approaches in this exciting and rapidly developing area of work. **Mechatronics '98** promises to be one of the largest specialist conferences in mechatronics held to-date world-wide.

147 high quality papers have been accepted by the International Programme Committee for presentation at the Conference in Sweden. They cover a wide range of mechatronics topics presented in the proceedings produced by Elsevier.

The conference programme covers a wide range of mechatronics related subjects which include; design methodologies and best practice; control and its application; industrial applications and manufacturing systems; non-industrial applications and case studies; robotics including mobile robotics and walking machines; novel sensors and actuators; vision systems in real-time control; virtual reality and disability; calibration, measurement and inspection; system modelling and simulation; and mechatronics education. Overall an extremely impressive collection of high quality contributions in related topics within the discipline of Mechatronics has been brought together.

This substantial international body of work, from authors in 34 countries is a showcase for the possibilities and benefits of mechatronics. We hope that you will find in these proceedings new ideas and new ways of applying established ones, which will be of use in creating innovative and valuable machines and systems for tomorrow. Furthermore, we hope they are a useful reference in facilitating mechatronics education.

Many people have worked to ensure the success of this event. Indeed, without teamwork and international co-operation it could not have happened at all. The Co-Chairs would particularly like to thank the members of the International Programme Committee for their work in refereeing all of the contributions and for providing, in many cases, invaluable suggestions to authors for improvements of clarity or focus. Special thanks are also due to the local Organising Committee members at the University of Skövde, for their guidance on conference direction and their practical local organisation. Besides the people on the Committee, Jeanette Karlsen the Conference Manager deserves a very special thank you all her hard work in managing and co-ordinating the preparation of most of the conference. Finally, thank you to the Institutions and commercial organisations who have sponsored this event by helping directly with the organisation or contributing to its organisational costs.

We hope that you thoroughly enjoy the conference and benefit from the whole experience and that we will meet again next time.

Paul Moore.

Philip Moore
De Montfort University

Hans Johansson
University of Skövde

Co-Chairs

Prof Philip Moore
De Montfort University, UK

Mr Hans Johasson
University of Skövde, Sweden

Organising Committee

Prof Philip Moore, De Montfort University, UK
Hans Johansson, University of Skövde, Sweden
Dr Memis Acar, Loughborough University, UK
Josef Adolfsson, University of Skövde, Sweden
Prof Sten Andler, University of Skövde, Sweden
Prof Alan Bradshaw, Lancaster University, UK
Dr Chris Czarnecki, De Montfort University, UK
Prof Abdulkadir Erden, Middle East Technical University, Turkey
Dr Patric Eriksson, Prosolvia Systems AB, Sweden
Prof Toshio Fukuda, Nagoya University, Japan
Prof Mats Hanson, Royal Institute of Technology, Sweden
Dr Rob Harrison, Loughborough University, UK
Dr Gert Johansson, University of Linköping, Sweden
Prof Oussama Khatib, University of Stanford, USA
Prof Tim King, University of Birmingham, UK
Prof Jeffrey Knight, De Montfort University, UK
Dr Lars Niklasson, University of Skövde, Sweden
Anders Nordqvist, University of Skövde, Sweden
Petter Olofsgård, University of Skövde, Sweden
Prof Rob Parkin, Loughborough University, UK
Jan-Christer Persson, IVF, Sweden
Prof Jury Podurajev, MSTU Stankin, Russia
Dr Jun Sheng Pu, De Montfort University, UK
Dr Klaus Selke, University of Hull, UK
Dr Paul Sharkey, University of Reading, UK
Prof Masayoshi Tomizuka, University of California, USA
Dr Mika Vainio, Helsinki University of Technology, Finland
Prof Gurvinder Virk, University of Portsmouth, UK
Prof Jan Wikander, Royal Institute of Technology, Sweden

International Programme Committee

AUSTRALIA
Prof John Billingsley, University of Southern Queensland
Dr John Mo, CSIRO

AUSTRIA
Dr Andreas Springer, Johannes Kepler University
Prof Peter Weiss, Johannes Kepler University

BRAZIL
Dr Marcio Luiz de Andrade Netto, UNICAMP-FEE
Prof Paulo Miyagi, Universidade de São Paulo
Dr Luiz G Trabasso, ITA-CTA

DENMARK
Dr Ole Ravn, Technical University of Denmark

FINLAND
Prof Aarne Halme, Helsinki University of Technology
Dr Mika Vainio, Helsinki University of Technology
Assoc Prof Tapio Virvalo, Tampere University of Technology
Prof Pentti Vähä, VTT

GERMANY
Prof Dr Ing Hartmut Janocha, Universität des Saarlandes
Prof Dr Gerd Hirzinger, DFVLR
Dr Friedrich Pfeiffer, Technical University of Munich
Prof Hubert Roth, Fachhochschule Ravensburg-Weingarten
Dr Stefan Rüde, Universität Karlsruhe
Prof Dr Klaus Schilling, Fachhochschule Ravensburg-Weingarten

GREECE
Prof Robert E King, National Technical University of Athens

HONG KONG
Prof Robin Bradbeer, City University of Hong Kong
Dr Ricky Yeung, City University of Hong Kong

HUNGARY
Dr Geza Haidegger, Technical University of Budapest
Dr O Petrik, Technical University of Budapest

ITALY
Prof Enrico Ravina, Universita di Genova

JAPAN
Prof Takashi Emura, University of Tohoku
Prof Toshio Fukuda, Nagoya University
Prof Iwao Hayashi, Tokyo Institute of Technology
Prof Yoshimi Ito, Tokyo Institute of Technology
Prof Masatake Shiraishi, Ibaraki University

MEXICO
Prof Victor Gonzalez, UNAM

NETHERLANDS
Prof Job van Amerongen, Twente University

NEW ZEALAND
Dr Reg Dunlop, University of Canterbury
Dr Saeid Nahavandi, Massey University

PORTUGAL
Prof Aníbal Traça de Almeida, University of Coimbra
Dr Hélder Araújo, University of Coimbra
Dr L M Camarinha-Matos, UNL Lisbon
Prof Carlos Couto, University of Minho
Dr Fernando Nunes Ferreira, University of Minho
Assoc Prof Mário Lima, University of Minho
Dr Júlio Martins, University of Minho
Prof João Luis Monteiro, University of Minho

RUSSIA
Dr Vladimir Pavlovsky,Russian Academy of Sciences
Prof Jury Podurajev, MSTU Stankin

SINGAPORE
Prof Poo Aun Neow, National University of Singapore
Tan Ah Sway, Ngee Ann Polytechnic

SLOVENIA
Asst Prof Matjaz Colnaric, University of Maribor

SWEDEN
Josef Adolfsson, University of Skövde
Prof Sten Andler, University of Skövde
Dr Patric Eriksson, Prosolvia Systems AB
Prof Mats Hanson, Royal Institute of Technology
Dr Gert Johansson, University of Linköping
Hans Johansson, University of Skövde
Dr Lars Niklasson, University of Skövde
Anders Nordqvist, University of Skövde
Petter Olofsgård, University of Skövde
Jan Oscarsson, University of Skövde
Jan-Christer Persson, IVF
Prof Bertil Svensson, Chalmers University of Technology
Sandor Ujvari, University of Skövde
Prof Jan Wikander, Royal Institute of Technology

SWITZERLAND
Prof Dr Ing Gerhard Schweitzer, ETH Centre

TURKEY
Prof Abdulkadir Erden, Middle East Technical University
Prof Dr Okyay Kaynak, Bogazici University

UNITED KINGDOM
Dr Memis Acar, Loughborough University
Dr Mark Atherton, South Bank University
Prof David Bradley, University of Abertay, Dundee, UK
Prof Alan Bradshaw, Lancaster University
Prof Cliff Burrows, University of Bath
Prof Jim Cooling, Loughborough University
Dr Chris Czarnecki, De Montfort University
Prof Peter Deasley, Cranfield Institute of Technology
Mr Ian Fraser, Royal Mail
Dr Mark Girolami, University of Paisley
Prof Roger M Goodall, Loughborough University
Prof J O Gray, University of Salford
Dr Rob Harrison, Loughborough University
Dr Mike Jackson, Loughborough University
Prof Tim King, University of Birmingham
Prof Jeffrey Knight, De Montfort University
Prof Gordon Mair, University of Strathclyde
Prof John S Milne, Dundee Institute of Technology
Prof Philip Moore, De Montfort University
Prof Graham Parker, University of Surrey
Prof Rob Parkin, Loughborough University
Dr Jun Sheng Pu, De Montfort University
Dr Geoff Roberts, University of Wales College, Newport
Dr Klaus Selke, University of Hull
Dr Paul Sharkey, University of Reading
Prof Gurvinder S Virk, University of Portsmouth
Prof Kevin Warwick, University of Reading
Prof Richard Weston, Loughborough University

UNITED STATES
Dr Dave Cliff, Massachusetts Institute of Technology
Dr Joseph Duffy, University of Florida
Prof Oussama Khatib, University of Stanford
Prof Eugene Rivin, Wayne State University
Prof David Russell, The Pennsylvania State University
Prof Masayoshi Tomizuka, University of California
Dr Jihong Wang, University of Texas Southwestern Medical Center

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Control of flexible assembly systems with a fault recovery capability

Yeung, W.H.R.[a] and Moore, P.R.[b]

[a]Department of Manufacturing Engineering and Engineering Management,
City University of Hong Kong,83 Tat Chee Avenue, Hong Kong.

[b]Mechatronics Research Group, Faculty of Computing Sciences and Engineering,
De Montfort University, Leicester, LE1 9BH, UK.

This paper presents a model which provides a fault recoverable environment for advance cell control in the application of Flexible Assembly Systems (FAS). The model integrates a flexible process planning (FPP) representation scheme, a conveyor based materials handling system (MHS), and an on-line production scheduler. The model relies on the ability to re-synthesise a Timed Colour Petri-Net (TCPN) cell controller when fault(s) occur. In the event of a fault condition, the current status of the manufacturing system is retrieved and used within the generation of a new production schedule. The ability to direct the re-synthesis of the cell controller enables the manufacturing system to adapt to the fluctuations imposed by conditions such as requests to add new products; changes in production volumes; the breakdown of assembly workstations / equipment; the unavailability of components; etc.. In this way the dynamic reconfiguration of the manufacturing facilities and operation plans utilised in the flexible assembly cell is made feasible within the constraints imposed by the production facilities.

1. FAULT TOLERANCE IN FLEXIBLE ASSEMBLY SYSTEMS

Fault tolerance is rapidly becoming one of the features to be embedded within modern highly automated intelligent standalone manufacturing machinery. Intelligent sensory devices such as force, torque, pressure sensors etc., together with the use of intelligent algorithms which allow real-time detection, diagnosis and even fault recovery are becoming more common. However, fault tolerant design in the domain of flexible manufacturing and assembly systems where usually a number of intelligent devices and materials handling systems co-exist and co-operate for the simultaneous manufacture of a mix of products is quite rare. The lack of fault tolerance in such systems often results in unacceptable performance in particular in unmanned systems where human operators are not readily available to diagnose and reset faults. This has contributed to the slow uptake of flexible manufacturing / assembly systems by industry.

Mendigutxia [1] presents a comprehensive model for a fault tolerant system in which he identifies the five functions: monitoring, prognosis, diagnosis, analysis of consequences, and programming of actions. He also outlines possible techniques including graphs rules, and Petri Nets, to realise these functions. Rezai [2] utilises a Petri Net approach to model faults in sensory systems and uses it for simulation and analysis. Jacot [3] and Wojcik [4] both attempt to design fault recovery strategies into Petri Net models for the real-time control of FMS. However, in both methods, additional global information about the manufacturing, such as the various process planning options, material control, and even faults states are modelled in the Petri-Nets before any control action takes place. This makes design of a Petri Net based controller very complex and some expertise in shopfloor control is also required. Moreover, encapsulation of too much information into the controller makes it difficult to expand and to interface with external components, such as scheduling systems, process planners, etc.. Here, however, a framework is implemented which attempts to provide a highly intuitive graphical interface to the user which facilitates the control of a general class of advance cell control system, which is capable of a level of self recovery in case of system faults. The framework is evaluated utilising a conveyor based flexible assembly system.

2. A GENERAL FAULT RECOVERY MODEL FOR ADVANCED CELL CONTROL SYSTEMS USING CONVEYOR BASED MATERIALS HANDLING

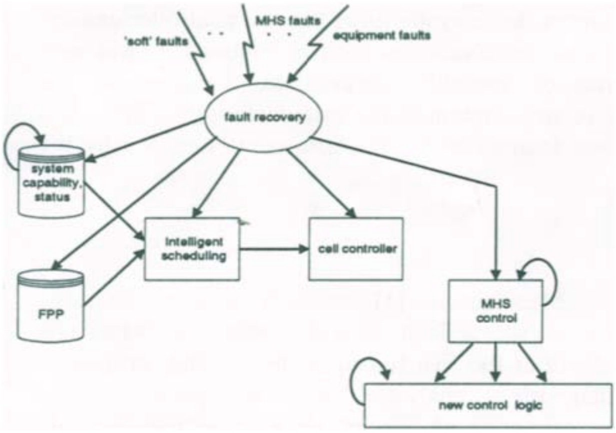


Figure 1. A general model for a fault recoverable cell control system.

Figure 1 illustrates a general framework for a fault recoverable cell control system. This framework consists of several autonomous and yet co-ordinated components, namely:

1. *"Update system capability and status"*. It is assumed that faults are identified and diagnosed from external processes and input into the model. As such, on receipt of a fault condition, there is likely to be a need to change the configuration or the operations of the manufacturing system, e.g. if a robot workstation has broken down, or a fault has occurred in the materials handling system. Furthermore, the model also supports business decisions which will sometimes necessitate the interruption and change of operation and process plans. In this case, we model such external influences as 'soft' faults and they are handled the same way as other 'hard' system faults.

2. *"Update the production schedule"*. When a fault exists, it is reasonable to assume that the system capability and status will be affected in some way. Therefore, there is a need to re-evaluate the production scheduling in particular the sequence of job orders, the selection of working operation plans (WOP's) and operations assignment accordingly.

3. *"Update the control logic of the cell controller"*. Most importantly, there is a need to update and re-synthesise the cell controller in order to synchronise with the new system capability and status. The model also accommodates the case that occurs in some cell controllers whereby the WOP of each product type are encoded within the cell control model. In this case, when a 'soft' fault is generated and it affects the selection of WOP, the cell controller will also refresh its control logic.

4. *"Update the materials handling control system"*. This model also supports advanced cell control systems which work co-operatively with materials handling systems of some sort. For instance, consider a conveyor based flexible assembly system, whereby faults may exist in the conveyor system which prohibit the availability of some routes. In this case, the control system for the materials handling system will update the physical configuration of the system and the control logic requirements.

3. IMPLEMENTATION OF THE FAULT RECOVERY SYSTEM

The previous section outlined the framework of a general fault recovery model for advance cell control. The mechanism for updating the production schedule was achieved by adopting a Genetic Algorithm to search in real-time for a sub-optimal schedule based on the latest system information [5].

This section reports on the mechanisms for updating the cell controller and the materials handling control system.

3.1. Synthesis of the CPN based cell controller

A basic requirement for a fault recovery cell control system is that it is able to adapt itself to synchronise with a new environment, namely, it can reflect the current manufacturing needs withstanding the constraints imposed by fault conditions. This involves the ability to synthesise the control model automatically when a fault condition occurs. In this model, a Timed Colour Petri Net (TCPN) is used as a cell controller for a manufacturing cell based upon a multi-robot facility and a flexible conveyor based materials handling system. The details of this cell controller can be found in references **[6]** and also outlined in appendix 1 below. The model of the TCPN structure necessitates that the following information has to be represented in the controller.

3.1.1. Information requirements and extraction

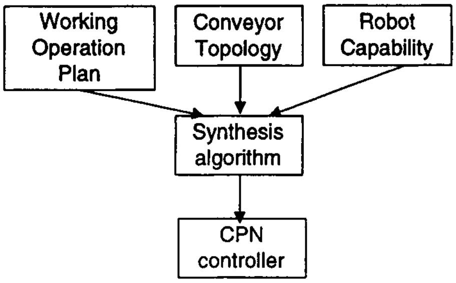


Figure 2. Information required for CPN synthesis.

As shown in figure 2, the Working Operation Plan (WOP) for each product i is; $WOP_i = [\ O_k \mid k = 1, \ldots, n_i]$ where WOP_i is a vector of n_i operations (O_k) to be exercised in sequential order; the capability of each robot server R_i at each workplace (WP_j) is denoted by; $RC_i(WP_j) = \{\ O_k \mid k = 1, \ldots, n_{Ri}\}$, which is a list of operations that robot R_i can perform at workplace WP_j; and lastly the conveyor topology, which describes the physical location of each workplace relative to each other, is; $CT = \text{Listof}\ \{WP_j, \text{Branch}_k\}$, where $\text{Branch}_k = \text{Listof}\ \{WP_j, \text{branch}_m\}$ and branch_k is the k_{th} branching point of the conveyor system. Although a CPN editor was developed to assist users to model the assembly system, it is still a complex and time consuming task to enter all the required information. Errors may be found even with experienced Petri Net users. In recent years, a number of methodologies have been proposed for the automatic synthesis of Petri Nets. Most of them are designed with Petri Net experts in mind who have a very good understanding of the modelling technique. The approach adopted here starts with a careful study of the information required. A graphical user interface (GUI) tool was developed to assist users on the shop-floor to input all the necessary information in a natural intuitive manner.

3.1.2. TCPN cell controller synthesis algorithm

The synthesis algorithm starts by scanning the conveyor topology list CT in a predefined order (as shown below). At each workplace WP_i retrieved from the topology list, the algorithm generates a set of operations ($Oset\_WP_i$) that can be performed. There after, it generates an array of all feasible operations ($Oset\_PD_j$) that can be processed for each product type PD_j. Note that at each workplace, there may be more than one robot that has operations at the workplace, and each of these robots may undertake more than one operation at the workplace. The algorithm compares the two sets of operations identified above for each robot and thus creates the CPN structure. The Pseudo English code in appendix 2 illustrates the algorithm used to synthesise the CPN controller automatically.

3.2. Fault simulator and controller for the flexible conveyor system

This module was developed to enhance the fault tolerance of flexible conveyor system (FCS) the details of which can be found in references **[7][8]**. As shown in figure 3, a simulator was designed which is responsible for sending artificial faults, both 'soft' and 'hard', to a fault controller. As highlighted in the general framework above, it is assumed that all faults have been identified and diagnosed when they enter the system, thus the severity level of the faults and the possible action that needs to be performed are known in advance. In the current implementation, the following faults can be simulated:

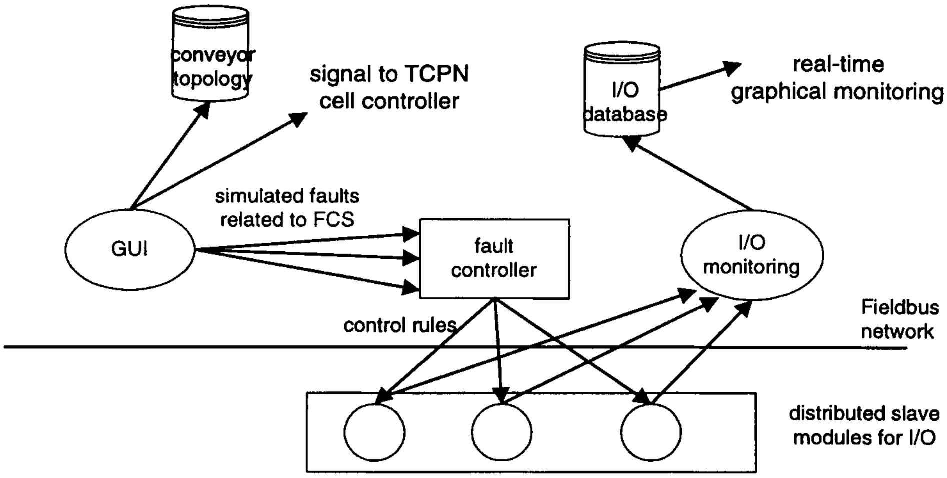

conveyor topology
signal to TCPN cell controller
I/O database
real-time graphical monitoring
simulated faults related to FCS
GUI
fault controller
I/O monitoring
Fieldbus network
control rules
distributed slave modules for I/O

Figure 3. Fault simulator and controller for FCS.

'Hard' faults:

1. *"Malfunction of a specific workstation"*. In this case, all products will bypass the workstation that has a breakdown fault condition. If the workstation is within a branch of a conveyor section, as shown in figure 4, then all products will be re-routed and do not enter this branch.

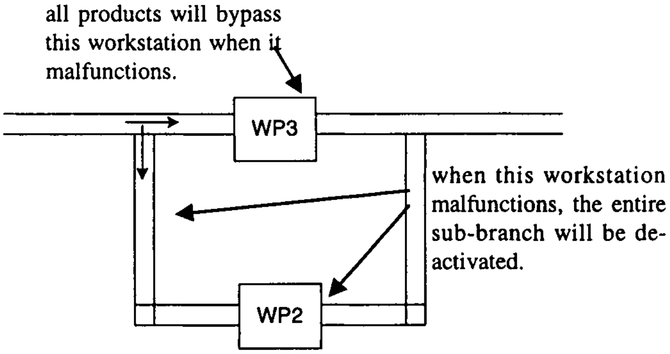

all products will bypass this workstation when it malfunctions.
WP3
when this workstation malfunctions, the entire sub-branch will be de-activated.
WP2

Figure 4. Possible actions for malfunction of a workstation

2. *"Failed operation at a particular workplace"*. In this case, the particular product will be directed through a pre-defined route to a manual repair station or reject station.

3. *"Failure of a specific I/O sensor"*. This is to simulate the situation where intelligent sensors [9] are employed which are capable of monitoring the reliability of the data they collect and generate alarm signals when the reliability level drops below some defined threshold. In this case, those control rules which use this sensor status will be de-activated until the sensor has been repaired or replaced.

4. *"Malfunction of the I/O ports of a slave firmware"*. This is to simulate the situation when one or more of the I/O ports of a slave firmware are likely to fail. In this case, external intervention is needed to resolve the fault and to inform the fault controller of any changes in the configuration.

'Soft faults':

5. *"Change the action of a specific branch point for a specific product type"*.

6. *"Change the routing of a specific product type"*. Both type 5 and type 6 fault simulate the situation where an arbitrary management decision is made to change the sequence with which a particular product type is manufactured.

7. *"Add routing information for a new product type"*. This is to simulate the situation when a new job order is added to the system whilst the system is operational.

8. *"Change the action of a specific branch point for all product types"*. This resembles the situation defined in the type 1 fault above.

These faults can be selected manually from the system. Once a particular fault type is selected, the system will send messages to a fault controller to inform it of the type of faults generated. In addition, when the faults selected will affect the overall topology configuration of the flexible conveyor system, for example, if type 1 and type 8 faults are selected, it changes the resulting topology of the FCS since one of the branches is in effect removed, the simulator will also update the data file of the conveyor topology and send a signal to trigger the automatic synthesis of the TCPN cell controller. The fault controller, when it receives the message from the fault simulator, will simply analyse the type of faults and generate the necessary control rules and download them for execution. The slave modules are so designed that they receive the control logic from the master node when they execute and continue to check for any messages from the master between each I/O scan cycle. A detailed description is presented in reference [7].

4. CONCLUSION

This paper proposes an embryonic framework for the implementation of a cell control system with a fault recovery capability. The proposed framework supports the integration of flexible process planning, intelligent scheduling and sequencing, and a flexible conveyor based materials handling system. A fault simulation system has been developed for evaluating the framework. The framework has been partially implemented to-date. In conclusion, the author's experience through the development of the systems indicates that the proposed framework is conceptually sound, and will provide a firm basis for further development of fault tolerant cell control systems.

REFERENCES

1. Mendigutxia J., Zubizarreta, P., Goenaga J.M., Berasategui L. and Manero L. "Fault Tolerance in Automated Manufacturing Systems", Expert Systems With Applications, Volume 8, No. 2, pp. 275-285, 1995.
2. Rezai M., Lawrence P.D. and Ito M.R., "Analysis of Faults in Hybrid Systems by Global Petri Nets", IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, Canada, Volume 3, pp. 2251-2256, 1995.
3. Jacot L., and Ladet P., "Combining the flow control techniques with high-level Petri Nets methodologies for real-time control of FMS", IFAC CIM in Process and Manufacturing Industries, Finland, pp. 133-138, 1992.
4. Wojcik R., Banaszak Z., and Roszkowska E., "Automation of self-recovery resource allocation procedures synthesis in FMS", IFAC CIM in Process and Manufacturing Industries, Finland, pp. 127-132, 1992.
5. Yeung W.H.R., and Moore P.R., "Advanced Cell Control in the Application of Flexible Assembly Systems", World Manufacturing Congress, Auckland, New Zealand, Nov., 1997.
6. Yeung W.H.R., and Moore P.R., "Distributed Cell Controllers using Colour Petri Net and Fieldbus: An application in Flexible Assembly", International Journal of Production Research, Volume35, No.2, pp.327-340, 1997.
7. Yeung W.H.R., and Moore P.R., "A Distributed Field Message Controller for Intelligent Conveyor Systems", IEEE International Conference on Systems, Man, and Cybernetics, Vancouver, Canada, Volume 1, pp. 109-113, 1995.
8. Yeung W.H.R., and Moore P.R., "An Integrated Object Oriented Environment for Rapid Programming and Control of Flexible Conveyor System", Proceedings, International Conference on Mechatronics'96 and Machine Vision in Practice, Minho, Portugal, Sept., 1996.
9. Henry M., 1995. Sensor validation and Fieldbus, Computing & Control Engineering Journal, December, 1995, pp.263-269.

APPENDIX 1 The Petri Net model for our prototype FAS

A Colour Petri Net (CPN) is a five-tuple (P,T,C, IN, OUT) where

P : {p1, ..., pn} is a finite set of n places,
T : {t1, ..., tm} is a finite set of m transitions,
C(p) and C(t) : are the sets of colours associated with place $p \in P$, and transition $t \in T$ correspondingly,
IN(p,t) : $C(p) \times C(t) \rightarrow N$ is an input function that defines the directed arcs from input place (p) to transition (t). N is the set of all non-negative integers which defines the weight of the arcs.

OUT(p,t) : $C(p) \times C(t) \rightarrow N$ is an output function that defines the directed arcs from transition (t) to output place (p). N is the set of all non-negative integers which defines the weight of the arcs.

In the design of this model, only one simple colour set is used to define the different type of product entering the system, i.e.,

C(p) = C(t) = {c1, c2, ..., cj} where ci denotes the colour token of the set which represents the product type i, in the system, for $i <= j$. and j is the maximum number of product types supported in the system. Therefore, the IN and OUT function can be represented by a simple 1xn vector of size n as follow:-

IN(p,t) = OUT(p,t) = [a1 a2 ... aj] where ai is a non-negative integer that denotes the weighting of the input/output arc with respect to product type i. J is the total number of product types supported in the model.

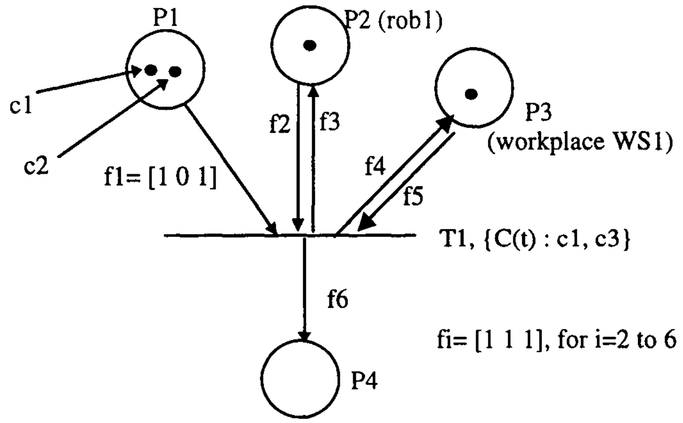


Figure 5 A sample Colour Petri Net model

Figure 5 shows a small section of the FAS model. The places, transitions and colours have the following interpretation.

P1 : input place where the arriving product flows through,
P2 : rob1 : availability of robot 1, a non-colour token at P2 denotes that there is one robot 1 available,
P3 : WS1 : availability of workplace 1, a non-colour token at P3 denotes that there is one workplace 1 available,
T1 : transition which represents an operation on an incoming product using rob1 at workplace WS1, the colour set C(t) denotes that this transition is only valid for product type 1 and 3,
C(P1) : colour set of place P1, and in this example, it denotes that there is one product of type 1 and one product of type 2 available.

fi : denotes the corresponding IN and OUT (1xn) vector function, for i = 1 to 6.

APPENDIX 2 Pseudo English code for the algorithm to synthesis the CPN controller

```
Struct Branch_element
{       int type; //branch point or workplace;
        int id;
}
CT_branch, CT is ListOf <Branch_element>;
// push the conveyor topology into a stack;
CT_branch stack = CT.Retrieve();
while ( Stack.Size() <> 0)
{   // pop the first element from the stack;
    Branch_element element = stack.GetFirstElement();
    if (element.type = "BRANCH")
    {   // push all elements within this branch into the stack, and continue
        element.Retrieve(stack);
        continue();
    }
    // else it is a workplace, and retrieve the workplace id;
    i = element.id;
    //Retrieve the list of operations Oset_WP(i)
    // System capacity is stored in relational database
    Embed_SQL(Select operation from system_capacity where workplace = i);
    while (NOT EOF())
        Oset_WP(i).append(operation);
    // Retrieve the list of operations Oset_PD(j) for each product
    for(j=1; j<=no_of_product; j++)
        for(k=1; k<=WOP(j).Size(); k++)
        {   Operation = WOP(j).FindElement(k);
            if (Operation.status = "COMPLETE")
                continue();
            // Compare WOP with Oset_WP to see if the operation can be performed at
            // this workplace
            if (Oset_WP(i).SearchElement(operation))
                Oset_PD(j).append(operation);
        }
    // Build the data structure of  the Petri Net building block for each robot which serves
    // at workplace i
    Obtain a robot_list which can serves at workplace i
    for (j = 1; j<= robot_list.Size(); j++)
    {   Robot = robot_list.FindElement(j);
        create a new timed transition;
        defined the corresponding input function of the input arc(by comparing
        Oset_WP(i) and Oset_PD(j) to determine which operations to be performed )
        construct the corresponding action list ( operations to be performed by which
        robot)
        set status of these operations to be "COMPLETE"
        if (Oset_PD(j).Size <> 0)
            continue();
        else
            break(); //go back to top of this algorithm for scan for next branch element
            // and  continue;
    }
```

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Development of an Open Embedded Machine Controller for the Automotive Industry

R Harrison[1] and L Lee[2]

[1]Department of Manufacturing Engineering, Loughborough University, Leicestershire, LE11 3TU, UK
[2]Ford Motor Company UK Limited, Dunton, Essex, SS15 6EE, UK

This paper focuses on the development of a modular open Embedded Machine Controller (EMCo) for the automotive industry. This system is currently being implemented through research at Loughborough University in collaboration with the In-Line and Diesel Engine (IL&DE) division of the Ford Motor Company UK Ltd., for use globally on future engine manufacturing programmes. The open control systems initiative OMAC (Open Architecture Machine Controller) instigated by GM, Ford and Chrysler in North America is reviewed and its suitability to meet the global manufacturing needs of the IL&DE manufacturing automation group at Ford, UK is evaluated. The limitations of vendor specific machine control systems and the motivation for adopting open approaches to machine control are considered.

1 INTRODUCTION

In manufacturing industry, shortening product life cycles combined with the need to offer increased product variety have made it necessary for machines to provide greater functionality, flexibility and reconfigurability. Much more is now demanded from control systems and as computer control systems have grown in complexity they have become much more difficult to design, build, operate, maintain, modify and integrate together. Most users to go to a single vendor for all their machine control needs in order to minimise such problems. The dominance of closed, vendor specific, solutions have resulted in stagnation rather than innovation and improvement in control systems [1].

The most common form of industrial controller is the PLC, accounting for more than seventy percent of sequence control systems installed in manufacturing industry today [2,3]. The PLC remains an essentially closed vendor specific product that has received little attention from the academic research community and its form still relates closely to its origin as a replacement for relay panels. There is now however an increasing trend towards the adoption of open modular architectures for this class of controller.

2 OPEN SYSTEMS

In recent years there has been increasing end-user awareness of open systems initiatives and standards. Their impact has been particularly evident in the domain of business and personal computer systems. Today many control system manufacturers are offering or announcing their new products as in some way being "open" systems. According to the IEEE definition "an open system provides capabilities that enable properly implemented applications to run on a variety of platforms from multiple vendors and inter-operate with other systems applications". In practise control systems are rarely implemented in an entirely open manner. The risks are generally too high for end users to make a step change to entirely open approaches. What is more likely is the progressive adoption of appropriate, available, proven and supported open systems.

Major motivations for the adoption of an open control system architecture are interoperability, portability, scalability and interchange ability. Interoperability enables system components to be designed independently and then integrated with components from *other* vendors. Portability allows system components to operate on different platforms. Scalability is a feature, which enables the customer to increase or decrease the functionality of a system by changing the number of components and/or upgrading or downgrading specific components. Interchangeablity allows substitution of one component with another due to its capabilities, reliability or performance.

3 OPEN CONTROL SYSTEMS FOR ENGINE MANUFACTURE

Within the IL&DE manufacturing automation group at Ford there has been an evolution of control system architectures from centralised to progressively more distributed forms. This trend can be seen in the design of transfer line machines for the machining of engine blocks. In the early 1980s these featured a centralised control system in very large overhead balcony mounted cabinets. Subsequent production lines have featured the distribution of the control onto each individual machine. The most recent engine lines employ a distributed control system architecture with vendor specific PLCs located on each station of the machine structure. Each PLC is responsible for the sequence control of actuators on its local machine station only. Ford now wishes to adopt a more open machine control approach on these production lines.

3.1 Open Control: Benefits and Concerns

From both end-user and machine builder perspectives open modular control systems promise reduced costs over a machine's life cycle [4,5]. Open, modular control systems, once established in the market could lead to:

- Reduction in development costs and in particular development time for highly automated manufacturing applications through a major change to the adoption of configurable, compatible automation components. These components being selected on a price/performance basis rather than vendor dependency.
- Reduction in the cost of eliminating faults (fast identification and replacement of defective building blocks) and ease of service and maintenance of the system components. Service and maintenance personnel being trained to support standard system components.
- Flexibility and adaptability of the system to changes in business direction during the life of the machine.
- Adaptations and alterations can be performed by the end user personnel themselves on the basis of an open modular system structure, without having to resort to an expensive system supplier.
- Breaking the dependency between machine controllers and their application software development. For example, allowing machine builders to re-use much of their generic software engineering and yet accommodate many of the end users' specific standards. (Machine tool builders should be able to efficiently move from a job in say Germany using company A's equipment, to another in North America using company B's equipment.)

There are however also concerns, from a user perspective, about the potential move to open systems and away from vendor dependant systems. Werner Harbers, head of Ford Motor Company's Manufacturing Research Centre has outlined reservations related to the fragmentation of responsibility for the integration of control system components, fragmentation of support after installation and also the risks of using a multitude of untried systems and suppliers [6].

3.2 OMAC White Paper

In December 1994 an "Open Modular Architecture Controller (OMAC)" white paper was published by a group of engineers from North America's three largest automotive producers. The document outlined their views on the requirements for open modular control systems in the automotive manufacturing industry [5] It also highlighted numerous problems related to the use of proprietary controllers that, while presented from an automotive viewpoint, are common to many other industries. Other publications related to OMAC have appeared more recently [7]. The control system requirements outlined in the OMAC report have been summarised and categorised into system requirements as follows:

- Meet all safety, reliability, robustness and environmental requirements.
- Satisfy government and company standards.
- Provide multi-vendor solutions.
- Easy upgrade without the involvement of controller suppliers.
- The use of components / tools available in the general-purpose computer industry.

- Intuitive and user friendly to reduce training expenditure.
- Support scalability by adding, removing or replacing modules.
- Present a common user interface environment across all applications.
- Be deterministic.
- Multi-level security access procedures.
- Provide a controller infrastructure that gives flexibility for integration of user proprietary technologies.

OMAC sub-system requirements are summarised in table 1.

Reviewing the OMAC white paper, in the context of the IL&DE automation group's needs for EMCo, the authors conclude the following:

- The scope of the EMCo application domain is narrower than that addressed by OMAC.
- OMAC aims to stimulate research and development projects and then utilise the best technologies and methods which emerge rather than providing a specification for next generation open control systems. EMCo aims to specify a controller architecture and to identify the interface requirements for sub-system modules; using where possible existing standards, specifications and products that are appropriate.
- OMAC appears to be primarily targeted at large centralised control systems. An EMCo platform is needed which is configurable to meet IL&DE's need for a new control system architecture to replace small modular PLCs which are currently used in distributed configurations.
- The OMAC document does not consider working practises and fails to address operational issues related to multi-skilling, which would greatly widen its appeal, particularly in Europe and Japan. This is seen as of key importance in order to ensure that EMCo can be globally supported and maintained.

3.3 Multi-Skilling Requirements for EMCo

A major factor effecting a control system over the bulk of its lifecycle is working practice. There are marked differences in working practises in different regions of the world (e.g. between North America, Europe and Japan). Open control system requirements specification and system design needs to consider the impact of such working practise if a globally acceptable controller architecture is to be created.

Sub-System	**Requirement**
External Interfaces	Allow connection to the upper business management layer using off the shelf components.
	Provide an interface to various real time bus systems (for the connection of inputs and outputs) with minimal cost and effort.
	Integrate diagnostic capabilities for both the controller and machine.
Internal Architecture	Provide flexibility to allow control designer to select the most appropriate system kernel for a particular application
	'Plug and Play' concept through the use a standardised Application programming Interface layer.
	Common real time system database.
	Controller hardware bus structure to be a 'de facto' standard, for example VME, ISA, EISA or PCI.
	The controller architecture must support standard output to servo drives, either digital or analogue.
	The control of the servo amplifier must be able to reside within the drive amplifier or OMAC.
Human Machine Interface (HMI)	Controller must support a commonly accepted graphical user interface environment, for example Microsoft Windows.
	The run-time version of the HMI must be available separately without the associated development system at a substantially lower cost.
	The HMI must have the ability to interface with other elements in the controller using a 'well accepted' messaging scheme, such as DDE in the Windows environment.
Application Software	Sequential control software to be written in IEC1131/3 or a flow chart style.
	The controller must support standard part programming inputs, such as RS-274D.
	The motion control editor should reside on the same platform as the HMI.
	The discrete event controller programming software will reside on the OMAC.

Table 1 OMAC sub-system requirements

Traditionally each shopfloor worker had considerable skill in a specific trade (e.g. electrician, mechanical fitter, machine operator). There was also typically strict demarcation between the tasks each tradesman undertook. In North America these traditional working practises still remain largely unchanged. In Europe and Japan however manufacturing plants have progressed toward multi-skilled operation [8]. This involves the use of small teams responsible for all aspects of production in an area.

When the programmable logic controller (PLC) was conceived the designers carefully matched the software and hardware to the existing skills of an electrician: as a result the PLC was extremely successful. However, many of the problems raised concerning contemporary use of PLC's reflect the fact that whilst the social and organisational techniques used on the shopfloor have changed extensively in the past twenty years; the PLC and its programming languages have remained fundamentally the same.

The design of the next generation of open control system must be undertaken with a complete awareness of this new multi-skilled environment in order to once again effectively re-match the skills available in the plant with those required to operate and maintain the control system. Consideration of multi-skilling gives rise to the following EMCo requirements (not addressed by OMAC):

- No repairs of hardware faults to be carried out on the machine. Design for rapid replacement.
- Simple reloading of application software without the need for any programming equipment.
- No specialised training required to use the control system. Less than 30 minutes training.

4 EMCo REALISATION

The implementation of the EMCo is based on an IBM Personal Computer (PC) platform and aims to utilise available infrastructural hardware components, software components, communication protocols and application development tools wherever possible.

4.1 PC Based Control

There have been many debates about the suitability of the PC platform for industrial control systems. Over the past decade however, the PC architecture has become an accepted platform for many control applications and its usage is expected to dramatically increase [9]. In the automotive manufacturing sector GM Power Train Group (GMPTG) has done a considerable amount to promote the use of open systems based on PC technology. In their recent projects in North America GMPTG have used in excess of two hundred 'open' P.C. based control systems.

The authors believe that a major factor now encouraging the more widespread industrial use of PC based control (which EMCo exploits) is the emergence of control network standards for motion and I/O. This has removed the problem of providing industrial standard I/O interfaces (a traditional strength of the PLC) on the PC since all I/O needs can now be met via 3 or 4 network connections (typically for motion, digital I/O, controller to controller and information systems communication)[1].

4.2 Packaging and Use of PC/104

The standard PC form factor is bulky and requires expensive mechanical packaging to ensure that a robust system is produced. Some of the OMAC white paper requirements are in practice difficult to achieve with the standard PC bus form factor and its associated card cages and backplanes. The use of the standard format would cause an OMAC to be larger than equivalent PLC systems; particularly on smaller machines giving little scope for cost saving. It is particularly inappropriate where a distributed control architecture is adopted requiring many compact, low cost controllers: an approach now being used on the majority of IL&DE manufacturing lines.

In the past the only practical way to embed the PC architecture in space and power sensitive applications has been to design PC functionality directly into the product. This however runs counter to OMAC's requirement for the use of 'off the shelf system components'. A more robust and compact implementation of the PC bus is therefore required without sacrificing hardware and software compatibility with the PC bus standard. PC/104 is a standard developed in response to this need. It offers full architecture, hardware and software compatibility with the PC bus, but in compact (3.6" x 3.8") stackable modules. In 1992, the IEEE started a project to standardise reduced form factor PCs. The PC/104 specification has been adopted as the 'base document' for this IEEE P966 standard. In September 1996 there was in excess of 150 different companies producing PC/104 boards [9]. In addition to the wide range of standard PC boards, (CPU's, Ethernet, serial communications, solid state memory etc.) a number of key suppliers in the industrial control field are producing PC104 boards. Boards are available for InterBus, Profibus and CAN (DeviceNet) and Indramat are about to launch a multi-axis motion control board using a SERCOS interface.

The EMCo is therefore PC/104 based and utilises a multi-layer carrier board. See figure 1. It is produced in two variants, which support either single or dual PC/104 systems. The PCs in the dual system communicate with each other via dual-port memory. The PC on the left-hand side of the carrier board (side elevation) can run MS Windows as the operating system, whilst the one on the right hand side can operate using a real-time kernel. In Europe where distributed control is commonly used it is expected that the vast majority of EMCo's will be based on a single PC only with an external display system attached via the fieldbus port.

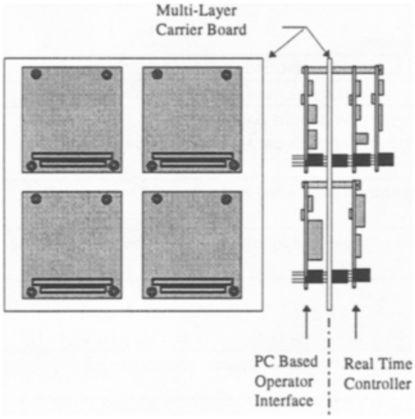

Multi-Layer
Carrier Board
PC Based
Operator
Interface
Real Time
Controller

Figure 1 Carrier board configuration

To enable the quick and easy replacement of EMCo units and the rapid reloading of application software by multi-skilled operators, application programs are stored on a PCMCIA card inserted into the side of the EMCo. When power is applied to the unit the application files are automatically distributed to the intelligent modules via the main processor module.

The approach to system software adopted on EMCo project is to single source the real-time operating system and the sequence control software. This ensures a single point of responsibility for this crucial area. The real-time operating system supports the necessary drivers for the interfaces to the other PC-104 modules used at run-time, typically for motion control, sensor/actuator bus and the safety module.

4.3 Architecture and Module Interconnection

The EMCo controller architecture is based a decomposition of the required system into standard control modules. The intention here is to provide are a set of encapsulated components with well-defined interfaces to meet the user requirement.

A major focus of the EMCo research is the creation of clearly defined hardware and software interfaces to ensure that the required modularity and scalability can be easily achieved. The aim is to minimise the software engineering effort that is required to initially integrate control components into a functional system, and also minimise the effort required to replace or upgrade portions of the control system when necessary [5].

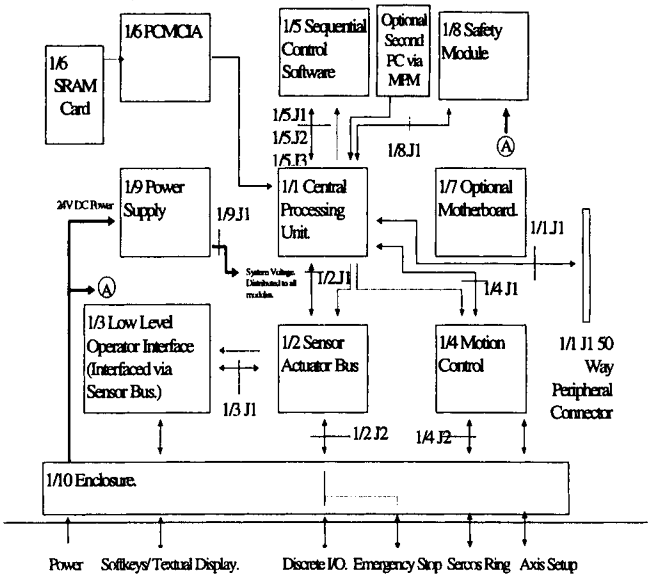

1/6 SRAM Card
1/6 PCMCIA
1/5 Sequential Control Software
Optional Second PC via MPM
1/8 Safety Module
1/5J1
1/5J2
1/5J3
1/8J1
Ⓐ
24V DC Power
1/9 Power Supply
1/9J1
1/1 Central Processing Unit.
1/7 Optional Motherboard.
1/1J1
System Voltage. Distributed to all modules.
1/2J1
1/4 J1
Ⓐ
1/3 Low Level Operator Interface (Interfaced via Sensor Bus.)
1/3 J1
1/2 Sensor Actuator Bus
1/4 Motion Control
1/1 J1 50 Way Peripheral Connector
1/2 J2
1/4 J2
1/10 Enclosure.
Power
Softkeys/Textual Display.
Discrete I/O.
Emergency Stop
Sercos Ring
Axis Setup

Figure 2 Controller architecture

Figure 2 identifies the major core modules required by the EMCo architecture (although clearly not all the possible configurations can be shown here). Each sub-system is given a number (e.g. *1/1 Central Processing Unit.*). A sub-system may be hardware, software or a combination of both. It will have one or more associated interface specifications. The notation used is *'sub-system number. interface number'(e.g. 1/4.J.1)*. Like the sub-system itself, the interface may be hardware, software or a combination of both. The model and associated specifications have been designed to allow specialist control suppliers to design and test products independently. Compliance with the specification requires that any sub-system can be added, removed or exchanged for another without effecting other systems. There are three parts to each sub-system specification currently under development: a section describing the 'hardware-system architecture', a second describing the 'instruction-set architecture' and finally a third, describing the testing and certification

Unit or Interface	Description
1/1 Central Processing Unit	The main processor must be compliant with the PC/104 V2.2 specification and offer a high degree of compatibility with the IBM PC family of computers. The compatibility must extend from the MSDOS-level, through BIOS-level to register-level compatibility.
1/1 Interface Req.	1/1.J1: Ensures sub-system compatibility with external interface connections, e.g. Serial ports, parallel ports etc.
1/2 Sensor / Actuator Module	This module acts as the controller board for the sensor / actuator bus. The general standard and protocol is defined by the individual bus associations. Card addressing is defined. Initially it is intended that three bus protocols be approved: InterBus, Profibus DP and DeviceNet. It should be noted that both a master and slave bus module (or combined) will be required to support different machine requirements.
1/2 Interface Requirements	1/2.J1: Defines the address area assignment of the PC I/O port. 1/2.J2: Sets the mechanical connection between the board and the enclosure.
1/3 Op. Interface	The operator interface must be compatible with the bus system used (i.e. InterBus, Profibus DP or DeviceNet).
1/3 Interface Req.:	1/3.J1: Operator interface protocols already exist and are laid down by the major bus associations.
1/4 Motion Control Module	The motion module receives instructions from the main sequential program via a multi port memory. The module uses these instructions to act upon its own independent motion program. Drive commands are issued via a serial bus to the drive amplifiers, (e.g. SERCOS protocol.).
1/4 Interface Requirements:	1/4.J1: Defines the address area assignment of the PC I/O port and function of individual registers. 1/4.J2: Serial Interface to Drive Systems.
1/5 Sequential Control Module	This is a software module containing the compiled application code and embedded operating system. The programming language and operating system are not defined.
1/5 Interface Requirements:	1/5.J1: Software interface to Sensor Actuator Module. 1/5.J2: Software Interface to Motion Control Module. 1/5.J3: Software Interface to Safety Module.
1/6 Memory Module	The memory module conforms to the PCMCIA standard. The SRAM card contains the application files for all modules. On power-up the files are distributed to the modules by the main CPU.
1/7 Optional. Motherboard.	Acts as a carrier for the PC104 cards if the self-stacking option is not used.
1/8 Safety Module	The safety module provides CPU scan monitoring and ensures 200ms-shutdown warning in the event of power failure. The power supply is designed to give 300ms hold-up of the 5v supply to ensure that the CPU can write the current I/O status table to the SRAM. External indication of scan and power supply failure is given via 'volt free' contacts.
1/8 Interface Req.:	1/8.J1: Defines the address area assignment and CPU interrupt for power and scan failure.
1/9 Power Supply Module	Provides 5v and 12v DC supply for all modules. The supply may be in a PC104 format (in smaller systems) or as a stand-alone device. Design criteria demands a 300ms hold-up of the internal supply in the event of an incoming supply failure.

Table 2 Modules and interfaces of the EMCo architecture

requirements. A full specification of each module is currently being developed however a partial description of each sub-system implementation and its interface requirements are given in table 2.

5 INTEGRATED USER ENVIRONMENT

Improving the integration of the application design/programming and the target control system environments is an important aim of EMCo. In most contemporary systems these environments are fragmented with little or no checking for data consistency. Typically there will be separate interfaces between each vendor specific software tool in the design/programming environment and each real-time sub-system (motion, sequence control, I/O, operator interface etc.). See figure 3. EMCo aims to provide a consistent windows based design environment for sequence, motion and operated interface aspects of application design. Application data is created and managed consistently and the target control system architecture integrates the real-time sub-systems. This approach allows a single PCMCIA card to hold all target system programs and data.

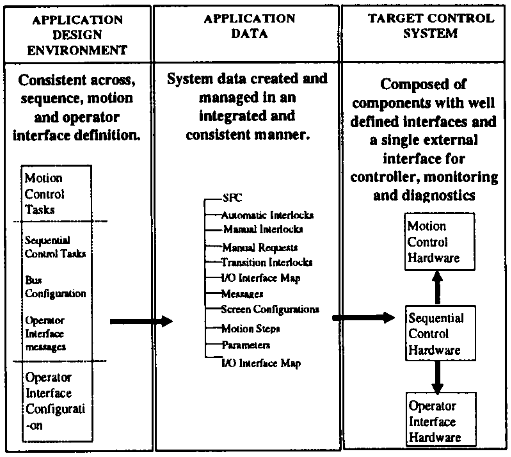

APPLICATION DESIGN ENVIRONMENT
APPLICATION DATA
TARGET CONTROL SYSTEM
Consistent across, sequence, motion and operator interface definition.
System data created and managed in an integrated and consistent manner.
Composed of components with well defined interfaces and a single external interface for controller, monitoring and diagnostics
Motion Control Tasks
Sequential Control Tasks
Bus Configuration
Operator Interface messages
Operator Interface Configurati-on
SFC
Automatic Interlocks
Manual Interlocks
Manual Requests
Transition Interlocks
I/O Interface Map
Messages
Screen Configurations
Motion Steps
Parameters
I/O Interface Map
Motion Control Hardware
Sequential Control Hardware
Operator Interface Hardware

Figure3 Integration of design and control

The application programming environment is known as MSFC or Machine Sequential Function Chart. It has some similarities to the SFC defined within the IEC1131 standard but is specifically targeted at efficiently supporting the programming, diagnostics and display requirements of the machine builder and user [10,11,12]. For example different types of steps are specifically supported for initialisation, wait, check and move operations. Also machine operational modes are inherently supported (for automatic, single step, manual and set-up operations). It is also important to note that definition of the diagnostics is an integral part of program definition.

Provision of an integrated approach to sub-system diagnostics is another important facet of EMCo. In the past each sub-system (motion, control, sequence control, I/O interfaces etc.) would have its own separate diagnostics system with its own proprietary interface.

6 CONCLUSIONS AND FUTURE WORK

The requirements for an open controller (EMCo) to meet the global needs of Ford IL&DE operations have been explained. The white paper on open modular architecture controllers (OMACs) issued by North America's three major control system vendors has been reviewed. The OMAC design requirements have been extended to take into account the needs of the multi-skilled workforce found in Europe and Japan.

EMCo is a planned progression from existing systems; it is "evolution not revolution". It embodies a more open component based approach, it supports distributed control of machine stations, it aims to offer a system that is more appropriate to today's operator/user skills set and it is designed to cope more easily with changing system configurations.

Future work will include a detailed evaluation of the EMCo system against current PLC based systems on both machining and assembly operations for identical use scenarios. This will include assessment of system flexibility, system agility, operational efficiency, scalability, availability, accuracy / performance, purchase and lifecycle costs.

ACKNOWLEDGEMENT

The support of the Ford Motor Company Limited is gratefully acknowledged in carrying out this research.

REFERENCES:

1 Lee L, "Next Generation Control Systems in the Automotive Industry", Proceedings of ICP International Conference, Paris, October 1996.
2 Anon, "The Future of Industrial Machine Control in Factory Automation", Industrial Controls Consulting, Inc., WI 54935, 1993.
3 Michel, G., "Programmable Logic Controllers: Architecture and Applications", John Wiley & Sons, Chichester, England, 1990.
4 Anon., 1994, "MOSAIC, Modular Open System Architecture for Industrial Motion Control", ESPRIT II project 5292.
5 Ford Motor Company, General Motors Corporation, Chrysler Corporation, Requirements of Open, Modular Architecture Controllers, Version 1.1, 1994.
6 Harbers W O, "Ford Automation Strategies and Needs", Automation Research Corporation: Automation Strategies Forum, Boston, Massachusetts, June 1996.
7 Anon., "OMAC? Oh Boy!", Carefree Communications Inc, Box 5268, Arizona 85377 USA, April 1995.
8 Anon., "Project Analysis Data: In Line and Diesel Engine Operations", Ford Motor Company, Dunton UK, 1994.
9 PC104 Consortium, What is PC104, 1994, PO Box 4303, Mountain View, CA 94040 U.S.
10 Anon.,"IEC Standard for Programmable Controllers, 1131 Pt 3: Programming Languages", Working Group 65A (Secretariat 67).
11 Bekkum, J., 1993, "The Coming of Open Programmable Controller Software", Control Engineering, October.
12 Anon, "MSFC Tutorial", Klopper und Wiege Software GmbH, D- 32657 Lemgo, Feb. 1997.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Petri Net-Based Inference Network for Design Automation under Nondeterminism Applied to Mechatronic Systems

Zühal Erden[a], Aydan M. Erkmen[b] and Abdülkadir Erden[c]

[a]Industrial Engineering Department, Çankaya University,
Öğretmenler Cad. 112/2, 06530 Balgat-Ankara, Turkey

[b]Electrical and Electronics Engineering Department , Middle East Technical University,
06531 Ankara, Turkey

[c]Mechanical Engineering Department, Middle East Technical University,
06531 Ankara, Turkey

This paper introduces the completed part of an ongoing research, in which a Petri Net-based design inference network is developed for the representation and analysis of the functions and their interrelationships through information flow for the conceptual design stage of mechatronic systems in order to facilitate design automation. The theoretical framework behind the network is based on transition of Hybrid Automata into Petri Nets and application of this framework is introduced by a mechatronic design example.

1. INTRODUCTION

The present research is directed towards automation of the mechatronic design process by modelling mechatronic integration at the functional design stage through a decentralised design inference network architecture [1]. Our aim is to develop a modelling framework for representing design alternatives on a *functional basis* and an evaluation with respect to model-dependent measures to obtain a rating for these alternatives. A mechatronic system has both discrete and continuous state changes and evolution. Therefore, it can be regarded as a hybrid system [2] for which a powerful modelling technique namely hybrid automata exists [2-5]. However, hybrid automata models focus mainly in representing *what* the states are and *what* are the conditions for switching from one state to another in terms of continuous and discrete behaviour. They do not provide on the other hand, explicit information about *how* this is done. Hybrid automata alone is useful when we try to analyse system behaviour. Our purpose in this research however, is to synthesise the system behaviour by examining *how* the required behaviour is obtained.

Petri Nets [6-9] allow to model dynamic systems where the information flow is important. Mechatronic systems are governed by information flow at the most abstract level, which corresponds to the conceptual design stage in a design procedure. By means of a network based on Petri Nets, we can determine *how* the system undergoes discrete and continuous state changes by examining the information flow switching logically through the network. Such an examination can be performed for several design alternatives on a functional basis using evaluation measures. Since mechatronic systems are considered as hybrid systems, we use hybrid automata as an intermediate model, which supports automatic transition from the functional representation to the network model ($SMDM_{net}$). In the theoretical framework, we also differentiate deterministic cases and nondeterministic cases, which involve uncertainties in information such as inconsistencies, incompleteness and imprecision.

This paper is structured as follows: The theoretical architecture for creating the network is presented in Section 2. Section 3 introduces uncertainties in information that may exists in the $SMDM_{net}$.

Application of the theoretical framework for creating the network is explained in Section 4 by using a mechatronic design example. Finally, discussion and conclusions regarding the presented study are given in Section 5.

2. THEORETICAL FRAMEWORK

The mechatronic system design presented in this paper, is based upon considering the system as being composed of multiple discrete configurations, in which the system behaves continuously within each one's constraints. Discrete jumps between configurations occur when constraints are violated. Thus, the theoretical framework of the network is initiated from a hybrid model of a functional design alternative that is further modified into a Petri Net-based network model. This framework forces the designer to define variables given in Section 2.1 for a functional design alternative.

2.1. Variable Representations

- *Functional State Set (FS)* is a finite set FS = { F_1, F_2, F_3, ... , F_N} of N sub functions (functional states) of the design alternative at the first level of functional decomposition [10].
- *Continuous Variable Set (CVS)* is a finite set CVS={V_j^C : j = 1, 2, ..., N_c } of N_c continuous variables which are the state variables defined for the unique representation of the continuous behaviour for the design alternative.
- *Functional State Matrix (FSM)* is an (NXL) matrix where L is the maximum possible size of the set of continuous functions for functional states.
- *Discrete Variable Set (DVS)* is a finite set DVS={V_k^D : k = 1, 2, ..., N_D } of N_D discrete variables which represent the types of configuration based information required to be perceived and controlled throughout the operation of the system.
- *Instantiations of the Discrete Variables (DVS_{ins})* are the binary values of the elements of DVS.
- *Invariant Conditions for the Functional State, F_i ($Inv(F_i)$)* are defined by a predicate over (DVS$\cup$CVS), describing constraints over the variables for the continuous behaviour in F_i.
- *Instantiations of the Continuous Variables (CVS_{ins})* are described by thresholding the elements of CVS according to the invariant conditions for functional states.
- *Discrete State Transition Matrix (DSTM)* is an (NXN) matrix, which represents the possible configuration-based discrete state transitions.

2.2. Creation of the SMDM$_{net}$

At this stage, the *Hybrid Model (H_F)* for the design alternative is automatically created as a 6-tuple;

H_F : (Loc, Var, Lab, Act, Inv, Edg)

In H_F, *Loc* is a finite set of vertices called *Locations*, which stand for the functional states of the system. *Var* is a finite set of real-valued *Variables*, which are defined as (DVS$\cup$CVS). *Lab* is a finite set of *Synchronisation Labels* used for the synchronisation of two hybrid automata. *Act* is a finite set of labelling functions that assigns to each location a set of *Activities (Act (F_i))* representing the continuous behaviour for the corresponding functional state. The labelling function *Inv*, defined by a predicate over *Var*, assigns to each location an *Invariant Condition (Inv (F_i))* such that *(Act (F_i))* holds when *(Inv (F_i))* is true. *Edg* is a finite set of edges called *Transitions*. Each transition e_{ij} is described as;

e_{ij} : (F_i, μ, F_j) where, F_i is the source location, F_j is the target location, μ is the transition relation (jump condition) which is defined as the predicate Inv (F_j).

The H_F is then modified into a Petri Net-based design inference network (SMDM$_{net}$), constructed automatically as a 4-tuple;

SMDM$_{net}$ = (P, T, I, O) where;

1) $P = P_1 \cup P_2$

$P_1 \subseteq (CVS_{ins} \cup DVS_{ins})$ and $P_2 \subseteq (CVS \cup DVS)$

2) $T = Loc \cup DF$

where DF = {df_1, df_2, df_3, ..., df_g}, g being the number of elements in P_2 and DF is a finite set of *Decision Functions (DFs)* which simulate the ability of a mechatronic system to decide how to process the incoming information. DFs are represented in the SMDM$_{net}$ as transitions in the form of *switches*.

3) $I : P X T \rightarrow \{0, 1\}$ is a mapping (defined by the binary values of 0 or 1) from places to transitions and can be represented by a predicate: I(p, t).

4) $O : T X P \rightarrow \{0, 1\}$ is a mapping (defined by the binary values of 0 or 1) from transitions to places and can be represented by : O(t, p) = I^{-1}(p, t).

A *marking M of the $SMDM_{net}$* is a g-dimensional column vector in which each element is another column vector of dimension (n_r +1), where n_r is the number of instantiations for each variable. A marking M therefore can be represented as;

$$M = \begin{bmatrix} \begin{bmatrix} m(p_{10}) \\ m(p_{11}) \\ . \\ . \\ m(p_{1n_1}) \end{bmatrix} & \begin{bmatrix} m(p_{20}) \\ m(p_{21}) \\ . \\ . \\ m(p_{2n_2}) \end{bmatrix} & . & . & \begin{bmatrix} m(p_{g0}) \\ m(p_{g1}) \\ . \\ . \\ m(p_{gn_g}) \end{bmatrix} \end{bmatrix}^T$$

where,

1) $m(p_{r0}$, r=1,2,..., g) is the number of tokens in a place (p_r) representing a variable (V_j^C, V_k^D),
2) $m(p_{r1})$, ..., $m(p_{rn_r})$ are the number of tokens in those places representing instantiations for the variable in p_r.

In the $SMDM_{net}$, there are two types of markings depending of the types of places in which tokens exist. These are a *variable marking (M^V)* and an *instantiation marking (M^I).* M^V represents a token state in which the system is represented by state variables (discrete and/or continuous) but it does not process them yet. M^V is a marking in which tokens are only placed in the variable places (represented by $m(p_{r0})$ where r = 1, 2, 3, ..., g). M^I is a marking in which tokens are only placed in the instantiations of variables (represented by $m(p_{rj})$ where r = 1, 2, 3, ..., g; j = 1, 2, 3, ..., n_r). M^I represents a token state in which the system obtains valid instantiations of the state variables by processing them and generating a decision through decision functions (DFs). An *initial marking M_0 of the $SMDM_{net}$* is a special variable marking *(M_0^V)*, which represents the initial information about the state variables of the system.

A marking $M_n^{V,I}$ of the $SMDM_{net}$ is said to be *reachable* from an initial marking M_0^V, if there exists a firing sequence which transforms M_0^V to $M_n^{V,I}$. A *firing or occurrence sequence* in the $SMDM_{net}$ is denoted by $\sigma_{SMDM} = M_0^V - DF - M_1^I - F_i - M_2^V - DF - M_3^I - F_i ..... - M_n^{V,I}$. In σ_{SMDM}, $M_n^{V,I}$ is reachable from M_0^V. The set of all possible markings reachable from M_0^V in a $SMDM_{net}$ is denoted by *$R_{SMDM}(M_0^V)$* and is called as the *reachability set of the $SMDM_{net}$*.

The dynamic behaviour of the $SMDM_{net}$ is based on the following assumptions : i) All of the enabled switch functions in the $SMDM_{net}$ (df_i, i = 1, 2,..., g) can fire simultaneously in a reachable marking $M_n^V \in R_{SMDM}(M_0^V)$ ii) Only one of the enabled transitions F_i, i = 1, 2, ..., N) in the $SMDM_{net}$ can fire in a reachable marking $M_n^I \in R_{SMDM}(M_0^V)$ iii) When a transition F_i fires in a reachable marking $M_n^I \in R_{SMDM}(M_0)$, any unused token in the places $p_{rj} \in P_1$ is reset. A $SMDM_{net}$ can be one of the two types as a *deterministic $SMDM_{net}$* or a *nondeterministic $SMDM_{net}$* with uncertainties. Definition of a deterministic $SMDM_{net}$ and its token flow is presented in [11]. In the following section, we present the definition of a nondeterministic $SMDM_{net}$.

3. NONDETERMINISTIC $SMDM_{net}$

A *nondeterministic $SMDM_{net}$* can be described as a $SMDM_{net}$ in which for $\forall M^{V,I} \in R_{SMDM}(M_0^V)$, $M^{V,I}$ involve uncertainties in information (signal) in the form of incompleteness, imprecision, inconsistency or noise.

Incompleteness in signal is defined as incomplete information in;

i) Places representing variables ($p_r \in P_2$),
ii) Places representing instantiations of variables ($p_{rj} \in P_1$)

and is represented by no token in some $p_r \in P_2$ in $M^V \in R_{SMDM}(M_0^V)$ for (i) and no token in $\forall\ p_{rj} \in P_1$ for some $p_r \in P_2$ in $M^I \in R_{SMDM}(M_0^V)$ for (ii). In this study, incompleteness is considered as a failure, which occurs mainly due to two reasons:

- Sensory Failure: Sensor(s) may be unable either to collect information from the environment or to send the acquired information to the system. This type of failure also covers environmental disturbances. That is, a sensory failure may occur either due to a physical failure within the sensor(s) or due to unexpected changes and disturbances in the environment. Sensory failure results in incompleteness of type (i).
- Process Failure: Any failure in decision functions (switch transitions) in the $SMDM_{net}$ may result in an incompleteness of type (ii). That is, the system obtains signal but is unable to send the processed output to the system as an

instantiation of a variable. Therefore, no token can be deposited in $\forall p_{rj} \in P_1$ for some places $p_r \in P_2$.

It should be noted that (i) results in (ii). Therefore, the problem of incompleteness in signal is considered as an instantiation marking $M^I \in R_{SMDM}(M_0^V)$ in which there exists no token in any of the instantiations of some variable places.

Imprecision in signal is a case in which the signal represented by a token in a place of the $SMDM_{net}$ has statistical existence. In other words, tokens exist in places with probabilities less than one. In case of *inconsistency in signal*, there exists tokens in more than one instantiation of a variable place, p_r. Each token has a probability less than or equal to one. State variables may further be contaminated by *noise*. Noise in a signal in the $SMDM_{net}$ decreases the probability of the existence of a token to a value less than one.

Definitions of uncertainty in signal introduced in this section results in the definition of a nondeterministic $SMDM_{net}$. Let $M^{V,I}$ be a marking of an $SMDM_{net}$ reachable from M_0^V ($M^{V,I} \in R_{SMDM}(M_0^V)$). $M^{I,V}$ is said to be a *nondeterministic marking (stochastic marking) of the $SMDM_{net}$*, if one of the uncertainty definitions above holds. Then, a *nondeterministic $SMDM_{net}$* is defined as a $SMDM_{net}$ such that for $\forall M^{V,I} \in R_{SMDM}(M_0^V)$, $M^{V,I}$ is a nondeterministic marking.

4. DESIGN EXAMPLE : A MECHATRONIC LACE CUTTING SYSTEM (MLCS)

In this section, we present a mechatronic design example for which a successful system was presented in [12]. The design problem is to precisely cut lace that moves on a conveyor. We present how a functional design alternative for a MLCS is described, using $SMDM_{net}$.

The proposed functional design divides the overall function of the MLCS into 4 sub functions at the first level of the functional decomposition as ;

- FS = $\{F_1, F_2, F_3, F_4\}$ with N = 4 and,

F_1 : MOVE LACE STRAIGHT,
F_2 : DETERMINE CUTTING PATH,
F_3 : CUT,
F_4 : STOP

- CVS depends on how a designer describes the continuous behaviour in each functional state. We take a general view as;

$CVS = \{V_1^C, V_2^C, \ldots, V_{NC}^C\}$

- We select a DVS composed of the following discrete variables:

DVS = {LM, CP, cutsit}, $N_D = 3$ where,

LM: a discrete variable representing the information about lace motion on the conveyor,

CP: a discrete variable representing the information about cutting path,

cutsit: a discrete variable representing the information about cutting situation.

- The following binary values are defined for the instantiations of the DVS;

LM = 1 (straight lace motion),
LM = 0 (lace motion is not straight),

CP = 1 (cutting path is determined),
CP = 0 (cutting path is not determined),

cutsit = 1 (cutting),
cutsit = 0 (not cutting).

- Inv(F_i) are defined as;

$Inv(F_1)$: $LM = 0 \wedge cutsit = 1$
$Inv(F_2)$: $LM = 1 \wedge CP = 0 \wedge cutsit = 1$
$Inv(F_3)$: $LM = 1 \wedge CP = 1 \wedge cutsit = 1$
$Inv(F_4)$: $cutsit = 0$

The Hybrid Model (H_F) developed for the MLCS is illustrated in Figure 1. H_F represents the continuous behaviour as functional states and conditions for allowable discrete transitions between functional states. $SMDM_{net}$ is then automatically generated using H_F and it is illustrated in Figure 2 with an initial marking. $SMDM_{net}$ describes *how* transitions occur within the system through information flow represented by token flow in the network, for which detailed formal definition is not covered in this

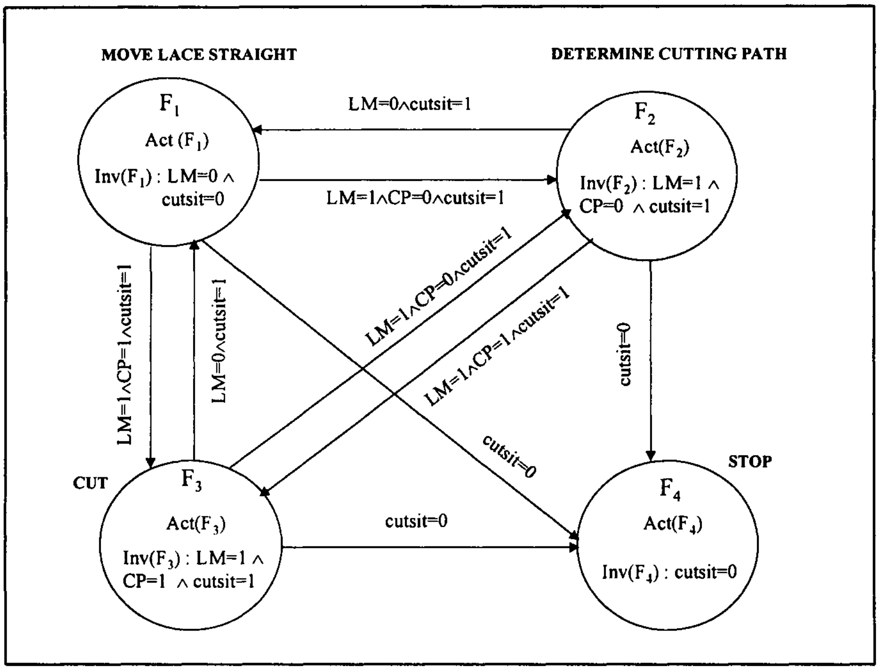

MOVE LACE STRAIGHT
DETERMINE CUTTING PATH
F1
Act (F1)
Inv(F1) : LM=0 ∧ cutsit=0
LM=0∧cutsit=1
LM=1∧CP=0∧cutsit=1
F2
Act(F2)
Inv(F2) : LM=1 ∧ CP=0 ∧ cutsit=1
LM=1∧CP=1∧cutsit=1
LM=0∧cutsit=1
LM=1∧CP=0∧cutsit=1
LM=1∧CP=1∧cutsit=1
cutsit=0
cutsit=0
CUT
F3
Act(F3)
Inv(F3) : LM=1 ∧ CP=1 ∧ cutsit=1
cutsit=0
STOP
F4
Act(F4)
Inv(F4) : cutsit=0

Figure 1. Hybrid Model (H_F) for the MLCS.

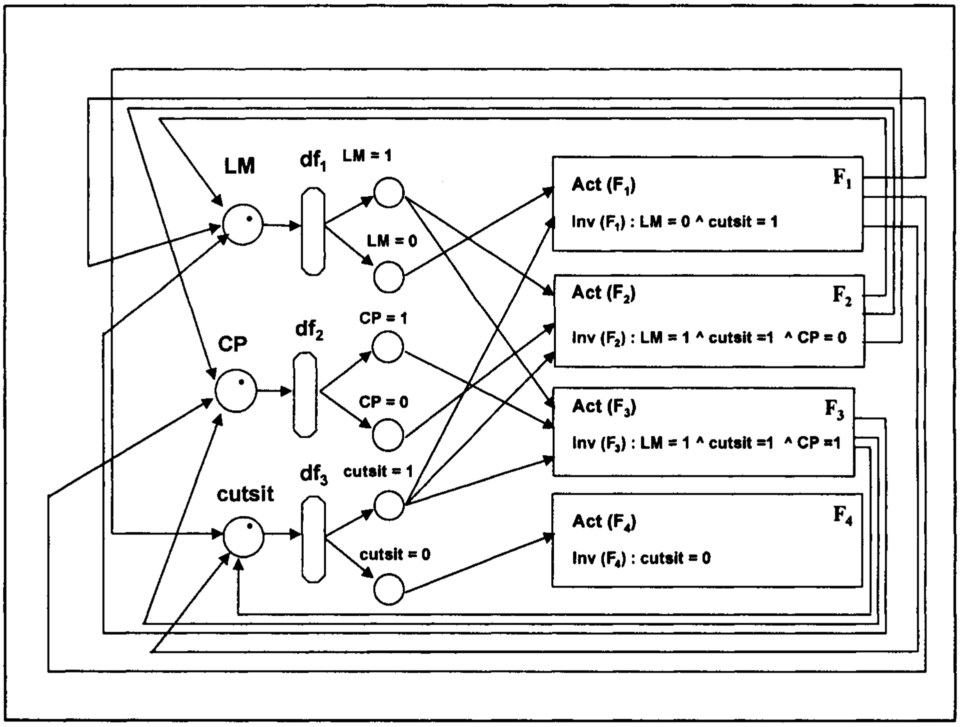

LM
df1
LM = 1
LM = 0
CP
df2
CP = 1
CP = 0
cutsit
df3
cutsit = 1
cutsit = 0
Act (F1)
F1
Inv (F1) : LM = 0 ^ cutsit = 1
Act (F2)
F2
Inv (F2) : LM = 1 ^ cutsit =1 ^ CP = 0
Act (F3)
F3
Inv (F3) : LM = 1 ^ cutsit =1 ^ CP =1
Act (F4)
F4
Inv (F4) : cutsit = 0

Figure 2. $SMDM_{net}$ for the MLCS with the initial marking (• represents tokens).

paper. SMDM_{net} provides the designer a graphical and algebraic tool for evaluating the dynamic behaviour of the design alternative.

5. DISCUSSION AND CONCLUSIONS

The present research aims to develop a design model applied to mechatronic systems, given a design need. In solving such a design problem, a designer is required to perform;

a) suggestion of several functional designs,
b) representation of suggested functional designs,
c) evaluation of suggested functional designs, and
d) physical realisation for to the functional design selected among successful functional designs with respect to predefined evaluation measures.

The design model developed in this ongoing research is a framework to automate step (b) and step (c) for mechatronic systems. These steps correspond to the conceptual design stages of a design procedure. The reason behind the use of Petri Nets is that; mechatronic systems are governed by information flow at the highest level of abstraction and Petri Nets are accepted as a powerful modelling approach for such systems. We work on a functional basis, which provides the network representation to be independent of any physical realisation and therefore, supports the inherent integrated characteristic of mechatronic system design.

Advantages of the presented network model can be summarised as follows: A realistic sub function in a SMDM_{net} can be used in another functional design (flexibility and multi resolution). Visual representation of created functional design alternatives is possible. Functional design is common to any engineering domain (SMDM_{net} supports integrated design).

We believe that, originality of the presented study resides mainly in the application of Petri Net theory in the conceptual design of mechatronic systems and second, in the integration of Hybrid Automata into Petri Net modelling, such that one can switch from "*what*" question to "*how*" question in the state space representation of dynamic behaviour of the system under consideration.

REFERENCES

1. Erden, Z., Erkmen, A. M. and Erden, A., "Modelling of the Job-Resource Interaction in a Mechatronic Design Process by Using Petri Net Theory'', Proc. 11th International Conf. On Engineering Design, ICED '97, Vol.2, pp. 753-756, Tampere, Finland, 1997.
2. Alur, R. et. al., ''The Algorithmic Analysis of Hybrid Systems'', Proc. 11th International Conf. on Analysis and Optimisation of DES, Lect. Notes in Control and Information Sciences 199, Springer-Verlag, pp.331-351, 1994.
3. Henziger, T. A. et al., "What's Decidable About Hybrid Automata?'', Proc. 27th Annual Symposium on the Theory of Computing, ACM Press, pp. 373-382, 1995.
4. Puri, A. and Varaiya, P., ''Verification of Hybrid Systems Using Abstractions'', Hybrid Systems II, LNCS 999, Springer-Verlag (1995).
5. Puri, A. and Varaiya, P., ''Decidability of Hybrid Systems with Rectangular Differential Inclusions'', Proc. 6th Workshop on Computer-Aided Verification, LNCS 818, Springer-Verlag, pp.95-104 (1994).
6. Murata, T., "Petri Nets : Properties, Analysis and Applications", Proceedings of the IEEE, Vol.77, No.4, pp.541-580 (1989).
7. Peterson, J. L., "Petri Nets", ACM Computing Surveys, Vol.9, No.3, pp.223-252 (1977).
8. Reisig, W., "Petri Nets : An Introduction", Springer-Verlag, Germany (1985).
9. Reisig, W., "A Primer in Petri Net Design", Springer-Verlag, Germany (1992).
10. Erden, Z., Erkmen, A. M. and Erden, A., "Generation of Functional Cells for a Mechatronic Design Network", Proc. The 5th UK Mechatronics Forum Int. Conf. (Mechatronics'96) with the 3rd Int. Conf. On Mechatronics and Machine Vision in Practice (M^2VIP'96), Vol.1, pp.233-238, Guimaraes, Portugal, 1996.
11. Erden, Z., Erkmen, A. M. and Erden, A., "Structuring of the: A Design Inference Network for Mechatronic Systems'' Proc. 2nd Int. Symposium on Intelligent Manufacturing Systems IMS'98, Sakarya, Turkey, 1998.
12. Preston, M.E. and King, T.G., "A Mechatronic Approach to Lace Scalloping", Mechatronics, The Basis for New Industrial Development, Ed. By Acar, M., et al., Computational Mechanics Publications, pp.493-499, UK, 1994.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Towards Paradigm Shift in Machine Design and Control

J Pu and P R Moore

Mechatronics Research Group
Faculty of Computer Science and Engineering
De Montfort University
The Gateway, Leicester LE1 9BH

This paper will present the author's thoughts and observations on recent changes and developments in machine design and control. Factors which cause such paradigm shifts will be identified. Areas of discussion will be related to PC-based control platforms, field-bus systems, machine design and simulation tools, smart products, components validation, machine life-cycle information and digital virtual engineering. The paper will present a components-based methodology, which is devised to meet such a paradigm shift in machine design and control. Key aspects of the proposed methodology will be outlined, e.g. views, images & diagrams; classification, objects and metaphors; physical platform (hardware), operating platform (software) and application platform ('virtual-ware'), etc.. The content of this paper forms the conceptual base for a recently launched ESPRIT project ('VIR-ENG'), which involves six industrial partners and two academic institutions. An outline description of the 'VIR-ENG' project will be presented, i.e. objectives, the methodology and test cases.

1. INTRODUCTION

European leading industrial sectors need to manufacture a wide variety of product models, which typically have a short product life, in short lead-times and in variable production volumes to retain competitive advantage in the global market place [1]. Even for the same model type, certain features variants are common to meet individual preferences. The underlying manufacturing strategy and systems must accommodate these constraints. The adoption of concurrent engineering and agile manufacturing has been advocated as a way forward in meeting such requirements. New concepts in building next generation production systems, which meet these demands, are now under development by major automotive manufacturers and associated partners. There is clearly a need for a coherent approach which embodies relevant perspectives in the manufacturing machine system life cycle to enable manufacturing solutions which withstand product variety,process variations, new product introduction and accommodate significant variations in production volume [2,3].

2. SHIFTS IN INDUSTRIAL MACHINE CONTROL PARADIGM: IMPACTS OF PCS AND FIELDBUS SYSTEMS

Figure 1 illustrates the authors' understanding of the current situation in process and machine control markets. PCs and fieldbus systems are seen as the two significant factors which can dictate the pattern of changes in various types of controllers. Nearly every PLC vendor now provides interfaces to PCs within their I/O networks [4]. PLC I/O systems are now under challenge by rapidly developing "third party" I/O and open field-bus systems. The vendors of field-bus systems also seek close relation with PCs. General-purpose digital motion control cards tend to be packaged with a user interface and hosted in the PC. PCs technology also stars to move into CNC machine and SCADA arena. Digital motion controllers are likely to provide some interfacing provision to some chosen field-bus protocol, either for loosely coupled applications or tightly coupled synchronization with usually small data transfer. To meet hard real-time requirements, backplane bus systems and/or real-time operating systems are still used.

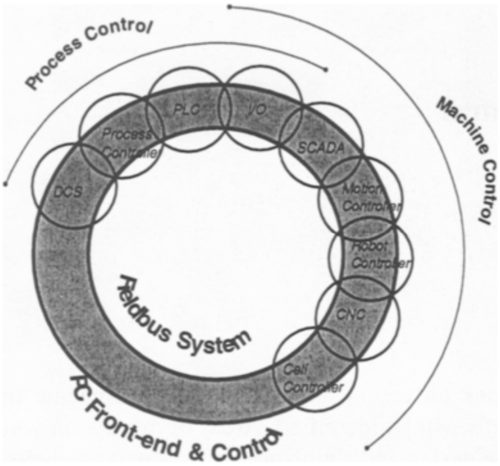


Figure 1 The Influence of PCs and Fieldbus Systems in Process and Machine Control Markets

2.1 Application of PCs in Industrial Control

The impact of PC technology e.g. Windows NT on industrial control is now widely recognized, which was perhaps not so obvious five years ago (gaining momentum from Windows 3.1) due to graphical user interface, hardware independence, multi-tasking and significant reduction in hardware cost. Initiatives on open architecture control systems have a lot to with PC-based control. PC are able to accommodate a control program which executes logic sequences and control loops without any specialist control hardware. Software logic and PC based control are the terms commonly used. PCs are able to run highly efficient object based application development tools.

Because real time control applications are typically time critical, open communications standards do not perform well for deterministic systems ("fast enough" is not adequate). When the requirement is for 'hard' real time control, even Windows NT has its limitations. After all, PC technology was not developed originally with industrial control in minds. PC products are engineered with a balance of cost and reliability to suit their office based, data processing primary application, strong in performing some form of supervisory control but weak in I/O capacity.

2.2 Fieldbus Systems

As machines become more sophisticated, distribution of intelligence is an important design philosophy for manufacturers and system builders to gain competitive advantage. Developments in computing electronics have resulted in fieldbus and control network technology which allow distributed control solutions to be realised cost-effectively and efficiently. 'Intelligence' can now be distributed to the component or device level, which provides a new dimension in building manufacturing machines and/or production systems. In recent years, fieldbus type technology has been widely applied in a number of industrial sectors, e.g. CANbus in some automotive designs and in manufacturing / machinery control where a number of application protocols have emerged including DeviceNet, and SDS; Profibus in the textile industry; LONWORKs [5] in building / home automation; P-Net and Interbus-S in process control; FIP in manufacturing; etc. The control elements can now be considered truly 'distributed' not only in conceptual terms but also in physical forms. As no single all prevailing international standard for fieldbus systems has emerged, then approaches which enable interoperability between dissimilar systems in a coherent and complementary manner, without limiting the inherent advantages of the individual bus systems, offer many attractions. Immediate and/or potential benefits identified are as the following

- controlling complex systems by using simple building blocks
- dramatic reduction in wiring and increased flexibility and modularity
- minimise machine's life cycle costs through improved fault self-diagnosis and lower downtime
- built on open industry standard for communications

However, significant design issues remain to be addressed, as such technology is still in its infancy (e.g. lack of user understanding, coherent development strategies and application tools).

3 SHIFTS IN DESIGN PARADIGM: INTERFACING/INTEGRATION BETWEEN "VIRTUAL" AND "REAL" WORLDS

The focus in the life cycle of machines and products is shifted more into the design phase [2,3], which to a large degree frames the demand for machine simulation & design tools. The requirements and benefits for virtual design or engineering have emerged [6]. The promotion of "digital plant technology" (PS-engine from Prosolvia Systems) can be seen as a recognition of this effect. Virtual engineering has a lot to do with virtual reality as well. However, the tools are not readily available for machine control and manufacturing applications.

Let agree that manufacturing systems or plant can be divided into four entities or subsystems, i.e. a) physical machines entity, b) actuators entity, c) controllers entity and d) sensors entity. With the simulation tools currently on the market, the modelling and simulation of controllers and sensors are not supported, which hinders two desirable attributes, i.e.

- The representation in the virtual worlds should be a true reflection of the real world.
- A seamless integration between the "virtual" and "real" worlds would be desired, although this may not be practical in many occasions.

As far as controllers and sensors simulation/emulation are concerned, the following functions could be developed to facilitate seamless integration.

- The application and/or control code developed in the 'virtual' environments can be ported to execute in the real-world run-time environment.
- The controllers can be used to drive the simulation ('virtual') system as well as the real-world system
- The sensors signal /data can be acquired from either the real-world or the virtual / emulation system.

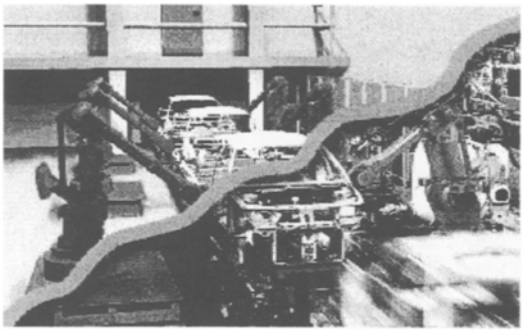


Figure 2 Seamless Integration between the "Virtual" and "Real" Worlds
(Courtesy of Prosovia Systems AB)

4 COMPONENTS-BASED SYSTEM ORIENTATION

Components-based view on machine design and control refers to the basic system orientation of machines and their control systems. This conception is devised based on the observations as follows.

- *Components or objects are the nature orientation of the universe.* This argument has been well explored in object-oriented software programming community.
- *Components-based software has been advocated as opposed to object-oriented programming.* An important argument is that object-oriented programming has not delivered what have been originally promised. An analogy can be made to what has happened to the flexibility expected from "robots". The most important attribute with components-based software is largely with the need for "encapsulation", whereby "inheritance" and "polymorphism" are not considered as essential.
- *Conventional machine design is largely components-based. In electronics engineering, the components-orientation nature can also be easily observed.*

However, somehow the components-orientation nature is not consciously observed and used in these days. The identification of components-based machine design and control is timely in the sense that

- The need for reconfigurable manufacturing systems is now well recognised in developed nations. Unless such systems are components-oriented, in both hardware and software terms, it is difficult to see how "reconfigurability" or flexibility can be achieved.

- Components-based orientation should be an effective and efficient way of tackling system complexity.
- Components-based orientation is a positive way forward in supporting "open systems" multi-vendor solutions.
- Developments in technology have enabled systems to be "connected" and "components-like" [5,7]. For example, networking is about connecting "components". Rapid development in field-bus systems in recent years have enabled "intelligence" to be extended to devices ("components") level (e.g. sensors, actuators, etc.) [8].
- The author(s) observes that the future trend in machine design and control is about "interfacing/connection" and "communication", in other words, "component-lised".

"Components-based" orientation is not a new thing. However, like many things, it is a conscious awareness and use of natural orientation as laid down from the beginning by the Creator.

5 A "3-D METHODOLOGY" FOR COMPONENTS-BASED MACHINE DESIGN AND CONTROL

To support components-based machine design and control, it would be helpful to have some level of systematic approach. This section presents some preliminary thoughts on what might be applicable. It can be seen as a collection, revising and orientation of relevant thoughts and observations. It is termed as a "3-D methodology" as "3-Dimension" seems to be a good way of describing and understanding the world around us. For example, a mechanical part can be fully described with 3-D or three 2-D drawing from three perspectives. Matters exist in three states "gaseous", "liquid" and "solid". Mechatronics is to study the combined effects of mechanical, electronics and computing software.

5.1 Principles of Similarity

- Classification & classes
- Objects
- Metaphors

Classification has been used by human beings since their existence for several thousand years, which powerfully helps them to understand the universe and the world around. According to its kinds ("classes"), objects are grouped. Metaphors can be seen as special objects, which help us to understand the particular nature of a "class" which the "metaphor" belongs and the behavior of similar objects which may be in entirely different forms. "Ability" can be seen in many ways the ability to apply what has been in one situation ("metaphor") to another situation. "Metaphors" have played an important role in human history of invention. Their importance can not be overly emphasized for design and learning.

5.2 Views, Images & Diagrams

For different observers, the "views" can be very different for the same "objects". From different "views", the "images" of the same object can be perceived differently. However, different "images" of the same object should be in conflict with each other. "Diagrams" are means of expressing the "images".

It is extremely important for the designer to be constantly aware of different "Views". Einstein's relativity theory is based on the observation from the "views" from two reference systems.

5.3 Components Types

Similar to the way matter is constructed, components may be seen in three different types.

- Primitive components – "atoms"
- Composite components – "molecules"
- Application components – "substances"

5.4 Three-Ware Model for Machine Design and Control Systems

- Physical Entity – Physical-*ware*
- Operating Infrastructure – *Software*
- Application – *Virtual-Ware*

The view in large for machine design and control systems can be encapsulated in the three-entity model, which is based on the following understanding.

- Reuse of application code, logic or programs, the virtual-ware, which is independent from the operating system or software infrastructure.
- The portability of the operating infrastructure (typically a operating system), which can be run on multi-vendor hardware platforms.
- The hardware platforms (the physical ware) are not restricted to a single source of supply.

The model supports multi-vendor open system solutions [9].

The proposition of the methodology is derived from Pu's earlier thoughts on ”the pervasiveness of similarity and modularity principles in 1988 during his period of PhD study.

6 THE ”VIR-ENG” PROJECT: INTEGRATED MACHINE DESIGN, CONTROL & SIMULATION OF MODULAR MANUFACTURING MACHINERY

6.1 Objectives of the Project and the Consortium

The project objective is to develop highly integrated design, simulation and distributed control environments for building agile modular manufacturing machine systems which offer the inherent capacity to allow rapid response to product model changes and feature variants. This acknowledges the need to compete effectively within a customer-driven market. The project seeks to achieve rapid introduction of new machine systems; rapid reconfiguration of existing machine systems when new products/models are to be introduced or process changes are to be made; simplified and proven retrofitting with minimum disruption to accommodate minor design changes in the product and quick change-over for predefined feature/model variants.

The consortium is comprised of a distinguished set of partners with complementary roles, which include end-users, machine builders / system integrators, CAD/CAE software developers, vendors of control systems and control components and science-base partners. That is *Volvo (project co-ordinator)*, a leading automotive manufacturer with a dual role as end-user and machine builder; *Krause,* the leading European automotive assembly/test systems builder/integrator; *Prosolvia,* a leading European developer and vendor of manufacturing systems and integration software; *Honeywell Controls*, a multi-national vendor of control and sensor technology; *Optimised Control,* an SME who is a leading vendor and innovator of intelligent motion control products; *Echelon*, a supplier of innovative control network technology and development tools; *De Montfort University* mechatronics research group with established expertise in machine control systems and the *University of Skovde* CIA research centre who specialise in graphical design and simulation of manufacturing automation.

6.2 Methodology for the Project

Three perspectives are identified to encapsulate the major attributes of the methodology, namely, the ‘design’ perspective, the ‘simulation’ (image) perspective and the ‘control’ perspective, as depicted in Figure 3.

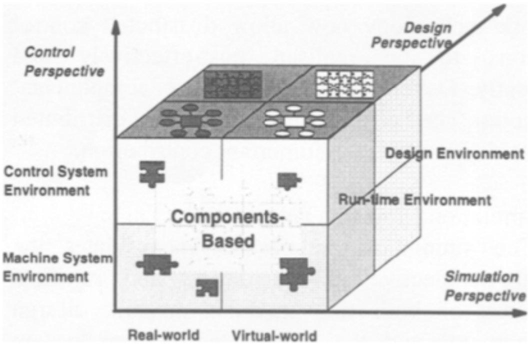


Figure 3 Reference Model of the Components-based Methodology for Machine Design and Control

- Control Perspective

The control perspective reveals the interaction between associated entities in a production system, viz. (i) modular machine design environment, to reflect the machine system and (ii) the distributed control system environment to reflect the control system. This perspective encapsulates the functional levels within a manufacturing machine, namely the ‘control system’ and the ‘machine system’. The machine system will be configured mainly from modular machinery elements, where these could include linear and rotary mechanical modules, conveyor systems, transfer units, etc.. The control system will be implemented in distributed forms as control components utilising recent developments in fieldbus technology, smart sensing techniques, and real-time operating systems. The control perspective is essentially component-based, in the sense that the overall machine system is built from the basis of ‘components’ which could, for example, be sensor components, control components and mechanical modules, etc.. Intelligence will be in-built into ‘components’ where appropriate to realise attributes of autonomy and cooperativeness. Components can be composed or aggregated together to form more complex components. These components can then be connected or bound together to form the specified machine system. The control hierarchy of the machine system will be decentralised. Once a machine or system has been configured, components can be replaced or updated. The supply of components, e.g. an intelligent sensor, a mechanical module, or a control unit, can potentially be sourced from multiple vendors. Developments in computing electronics and their peripherals, fieldbus and network technology now allow distributed control solutions to be realised cost-effectively and efficiently. Development of specialist 'components' for motion control and smart sensing in distributed systems is seen to be an important contribution.

- Simulation ('image') Perspective

The simulation perspective encapsulates the virtual perspective of the design and run-time processes of both the 'modular machine design environment' and the 'distributed control system environment'. The approach thus adopted for building distributed systems and processes is expected to foster one important feature, i.e. the creation of 'virtual' process and components, where 'virtual' is used in the sense that they do not physically exist but exhibit the 'image' of its physical existence and operations involved from the view of related components. This is realised through the separation of two basic properties in a 'component', i.e. the 'image' and its 'substance' (the repository in which the 'image' is formed) [10]. The 'image' of the virtual component is seen as the same for that of the real component. However, the substance of the 'image' can be manipulated which marks the difference between a 'virtual' and a 'real' component. The 'components' can be designed in such a way that when the substance of the component is changed the external view or 'image' of the component remains unchanged. Through the creation of virtual processes and components in both the design and run-time environments, it will enable concurrent design processes to be established and coherently integrated and on-line reconfiguration of the machine/system to be carried out, resulting in reduced lead time, machine downtime and product-changeover time. It is anticipated that a significant step change can be made in commissioning, installation, fault-diagnostics and testing of manufacturing machines.

- Design Perspective

Methods and tools to enable the 'virtual design / engineering' of machine systems is seen to a significant feature of the methodology. It is envisaged that the 'modular machine design environment' and the 'distributed control system environment' will be closely mapped, allowing a high level of integration based upon the underlying component based architectures. Machines systems have undoubtedly become more software-driven (or featured) and this has had a profound effect on the machine life cycle and the design processes involved. Machine system designs tend to be more dynamic, requiring the provision of building blocks with accompanying tools and methods for aggregating these building blocks. The design perspective encapsulates the design and run-time attributes of developing manufacturing machines systems. Previous research has confirmed the importance of a close mapping between the design and runtime environments. However, the emphasis of previous work has been to derive generic software code to allow vendor-independent solutions. In reality, much effort was spent in wrestling with implementation details of the supporting platform (e.g. operating systems). The strategy within the proposed project is that the runtime environment incorporates a level of design capability and the design environment similarly will embody run-time emulation attributes. A close interaction between the design and runtime environment extends to the component level in the system hierarchy. Visualisation of machine system 'concepts' and enactment of control logic and application logic in a virtual environment are important aspects of the approach. Diagnostic functions and life-cycle data acquisition and analysis mechanisms will be implemented where appropriate at both components and machine levels. With such provisions, warning signals and messages can be given to users when components or devices approach the end of their operating life. Patterns-of-use of machines will be recorded so that the history of operation can be checked when a machine break down occurs to help the operator or maintenance engineers to establish the source of the problem. Display panels and in particular PC-based window graphics will be utilised to facilitate visualisation requirements. Special debugging and diagnostic tools or devices may also be developed. The creation of diagnostic functions and life-cycle data acquisition and analysis is seen as a perspective of the human-centred approach. This is regarded as an essential requirement for software and/or electronics based systems, where the majority of efforts is often spent in identifying the problems when failures occur in contrast to mechanical systems whereby faults are relatively easy to identify but difficult to repair.

The development strategy for the VIR-ENG project is shown in Figure 4.

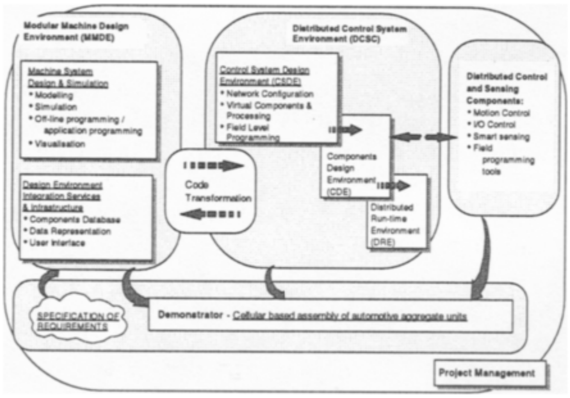

Modular Machine Design Environment (MMDE)
Machine System Design & Simulation
• Modelling
• Simulation
• Off-line programming / application programming
• Visualisation
Design Environment Integration Services & Infrastructure
• Components Database
• Data Representation
• User Interface
Code Transformation
Distributed Control System Environment (DCSC)
Control System Design Environment (CSDE)
• Network Configuration
• Virtual Components & Processing
• Field Level Programming
Components Design Environment (CDE)
Distributed Run-time Environment (DRE)
Distributed Control and Sensing Components:
• Motion Control
• I/O Control
• Smart sensing
• Field programming tools
SPECIFICATION OF REQUIREMENTS
Demonstrator - Cellular based assembly of automotive aggregate units
Project Management

Figure 4 Development Strategy for the VIR-ENG Project

6.3 Test Cases

The target application for the demonstrator is the flexible assembly of automotive aggregate units up to some 120kg. (e.g. small engines, cylinder heads, transmissions, axle components, steering gearboxes etc.). The demonstrator cell is likely to include modular machine elements, modular accumulating roller conveyor elements with assembly pallets, automatic stations for flexible assembly operations, manual stations, verification stations and production management information systems. The target production range for the cell would be between 50 to 2000 units depending upon the number of operations. A floor space of some 200 square metres is likely to be needed for this type of cell. This initial design brief for the cell has been chosen after careful analysis of the future manufacturing requirements in the European automotive assembly and test sector. A cellular manufacturing concept has been deliberately chosen.

Special features of the demonstrator cell proposed include:

- demonstration of 'agile modular machine systems' used to produce a real pilot production facility.
- state-of-art 'cellular manufacturing' principles are applied (process sequence layout).
- generic cell concept - to accommodate a wide range of automotive product types. The same basic cell structure and components could be used for different product types.
- the cell will be implemented using 'virtual engineering' methods and tools which support 'modular manufacturing machine system' design (visualisation, design, simulation, analysis and control code generation).
- the approach integrates automatic and manual stations in a cellular structure.
- the cell will be extendible and reconfigurable (in both a physical and a software sense).
- modular machine components (e.g. mechanical modules, sensors, control units, etc.).
- distributed control system with intelligent / smart components with inherent diagnostic capabilities.
- human centred approach adopted in design methods and implementation.
- skilled operators within the cell take on maintenance functions (exploiting the 'smart diagnostics' of the control system).

7 CONCLUSION

We may conclude that a paradigm shift in machine design and control is required towards a more distributed components-based configurable orientation in both software and hardware, with close interaction and mapping between the 'virtual' and 'real' worlds for both machine systems and their control system. The VIR-ENG project can make a lead in advancing this paradigm migration, through which a methodology in supporting components-based machine and control can be developed. It is anticipated that an extension to UML could prove to be a useful tool to support such a methodology. A new horizon in machine design and control should emerge through the application and integration of new enabling technologies and support tools.

ACKNOWLEDGEMENT

The authors would like to express their appreciation of the discussions made with their colleague Mr. C B Wong, which have proven to be so useful and beneficial. The authors also would like to acknowledge our gratitude to the consortium for their enthusiastic participation and the financial support of VIR-ENG project from European Commission.

REFERENCES:

1 Anon (1995), "CEDAM, a survival concept for the European machine tool industry". MB Produktietechniek, Vol. 61, No. 4, April, pp 104-109.

2 Weston R. H., et al, "A New Approach to Machine Control", Computer Aided Engineering Journal, February 1989.

3 Harrison R. et al, "Integrated Machine Design and Control (IMDC)", EPSRC/CDP Group, GR/J/57827, Loughborough University of Technology.

4 David Folley, "DCS or PLC? Or Perhaps just PCs", Control Systems, March 1998, pp17-18

5 Blomseth R and Capolongo W., "The LonWorks Component Architecture - A Software Component Architecture for LONWORKS Tools", Echelon Corporation, CA, USA, 1995, pp 323-352.

6 Eriksson P T and Moore P R, "Improved Computer Aided Robotics with Simulation of Sensors", Proceedings Mechatronics and Machine Vision in Practice 1994, IEEE Computer Society Press, ISBN 0-8186-6300-6

7 H Kopetz, "Component-based design of large distributed real-time systems", Journal of Control Engineering Practice, April 1998, Vol 6/1, pp. 53-60

8 Norris D., "Smart Distributed Systems: Distributed Control for Factory Floor Automation", FieldComms'95, Nov.7-8, Hinckley Island Hotel, Leicester UK

9 Harris, C.J. and Fraser, R., "Command and Control Infrastructures: the need for open systems solutions", Advanced Systems Research Group, University of Southampton.

10 Pu J. and Moore P. R., "Component/Image-based Design of Distributed Manufacturing Machine Control Systems", ICRAM'95, Istanbul, Turkey, August, 1995

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Modeling of failure treatment in manufacturing systems through Petri nets[*]

P.E. Miyagi, L.A.M. Riascos, N. Maruyama, F.G. Cozman

Escola Politecnica, University of Sao Paulo
Av. Prof. Mello Moraes 2231, CEP 05508-900, Sao Paulo, SP, Brazil

In this research, a variation of Petri nets called Self-Modifying Petri nets is applied for modeling failure treatments in manufacturing systems. The failure detection and treatment problem is discussed and the main concepts of Self-Modifying Petri nets are presented for modeling manufacturing systems. An example of a manufacturing system is presented to illustrate the effectiveness of modeling through Self-Modifying Petri nets. The proposed approach facilitates the analysis of the modeled system when compared to other types of nets especially in the case of discrete event dynamic systems with structural changes.

1. INTRODUCTION

The growing complexity of the structure and functional features of the Mechatronics approach for manufacturing systems demands the utilization of novel modeling and analysis techniques. The thechniques aim at both the optimal design alternatives and operational policies for increasing productivity/quality and reducing the manufacturing cost. Examples of such systems are the Anthropocentric Production Systems (APS) [1,2] and the Balanced Automation Systems (BAS) [3], where the interaction with the human operators evidences the need for strategies of treatment of failures and errors.

These systems can be studied from the point of view of Discrete Event Dynamic Systems (DEDS) [4-6] and Petri Nets (PN) [4-8]. These systems can be approached as a class of DEDS because their dynamic behavior are characterized by states that change only at discrete instants of time in response to the occurrence of specific events (events such as: message or part received, task completed, machine failed, process activated, etc.).

Based on the potential of PN, this work introduces the self-modifying Petri nets [9] technique for modeling failure treatment in manufacturing systems.

2. MANUFACTURING SYSTEMS, FAILURE DETECTION AND ITS TREATMENT

2.1 Modern manufacturing systems and failures

A modern manufacturing system [10] is a highly automated production system consisting of a number of computer controlled machine-tools and/or manufacturing cells that are interconnected for information interchange (by a supervision system) and material transfer (by a transportation system, composed by robots, AGVs, belt conveyors, etc.). The flexibility of this system depends on the diversity of the fabricated product (product flexibility), the adaptability of machines (equipment flexibility) and the similarity of processes (processes flexibility).

An APS [1,2] is a concept that can be defined as a human centered manufacturing system, as opposed to a technology centered system. This system is designed to increase the participation of people in the decision-making process and in the control of production process, thus leading to a better quality of working life (operator flexibility).

A BAS [3] is a generalization of APS, characterized by an adequate combination of automated tasks and human activities to obtain the highest autonomy and work level using limited technology and human resources.

---

[*] This research is partially supported by the Brazilian agencies: CNPq and CAPES.

Therefore, despite the advances in Mechatronics and related technologies, failure treatment in manufacturing systems is a fundamental problem to be solved because it is impossible to ensure that a design is error-free, or a equipment never fails, or human operators are perfect and never make mistakes.

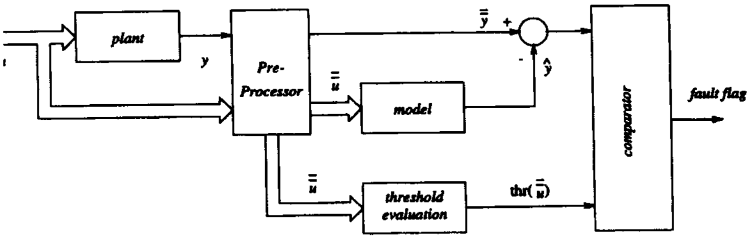


Figure 1. A failure detection scheme with static neuro-fuzzy model.

2.2. Detection of failures

It is possible to consider two level of failures:

- in the level of equipment elements there are failures in operator commanding elements (such as push-buttons), operator monitoring elements (such as lamps), sensor elements (such as level-switches) or actuator elements (such as valves),
- in the level of decision there are failures in the control/supervision of equipment and its elements (such as computer-hardware devices or software packages).

Failures of the first level are directly detected by monitoring the elements while in the second level, some methodologies for the identification of failures are necessary.

The strategies for detection and isolation of failures can be classified in two approaches [11]:

- the analytical redundancy approach: where the detection and identification of failures in the system is realized by comparing the data collected from the system with the data of functional models;
- the knowledge-based approach: where qualitative models associated with heuristic reasoning are applied for detection and identification of failures.

In many practical situations, a combination of both the analytical and knowledge-based methodologies should be the most appropriate solution. Figure 1 shows an approach presented in [12] where static neuro-fuzzy models are used to describe the correct operation of the plant. Abnormal operation of the plant is detected when residuals exceed a variable threshold which accounts for modeling errors, noisy measurements and unmodeled disturbances. Neuro-fuzzy models confirm its effectiveness for systems with high uncertainties specially when physical or semi-physical models are difficult to be obtained. Due to its quantitative and qualitative interpretation both training data and domain knowledge can be combined when building the model.

In general, decision-making techniques can be employed to detect and isolate failures. The complexity of manufacturing systems calls for techniques that can represent and manipulate large number of variables, such as Bayesian networks [13, 14]. These graphical structures have been proven to compactly represent complex systems and to produce efficient calculations of probability and expected values [13].

2.3. Failure treatment

There are two approaches for failure or error treatment in any equipment or system:

- non-autonomous systems: where the system only informs whether any component (or sub-system) does not execute some of their functions. In this case, generally, human procedures and interventions are defined for repairing the failure and the solution of the problem is an adequate combination of material and human resources [1-3];
- autonomous systems: where its components should execute the following functions [15]:
  - monitoring of system states;
  - detection of failures;
  - diagnosis of failures;
  - planning of repairs;
  - execution of repairs.

In this context, there are two strategies for executing the repair:

- the adjustment of parameters without changing or reorganizing the logical structure of the system, by controlling only the actuators to recover (totally or partially) the requested functions [15]. The execution of repair depends on the functional characteristics of the system; the functional redundancy, which changes the original structure of the system. This is achieved by including for example, additional redundant parts in the system.

3. PETRI NETS

The approaches for failure detection and treatment in manufacturing systems presented earlier define important concepts and procedures but does not explicitly show how they are effectively implemented in a system.

On the other hand, Petri nets (PN) have been proven to be a powerful approach for modeling and analyzing DEDS such as a manufacturing systems [4-8]. The PN is a technique that provides a uniform environment for representation, qualitative and quantitative analysis, and design of manufacturing systems.

Despite the PN has been utilized by various researchers in studies of manufacturing system, the emphasis was placed on modeling, analysis of performance and systems control under normal operations and normal conditions. Few studies have been conducted to apply the PN theory to the problem of failure treatment.

3.1. Modeling and treatment of failures

The failure treatments can be modeled by sub-nets. In these cases, depending on the type of the failure, four approaches can be considered [8]:

- the input conditioning method: in this case if a failure state is detected in a *place* of the net, it can become a normal state after the execution of a correction action in the system. Figure 2 illustrates the failure treatment through this method, where Z is the original PN and S is a sub-net representing the failure treatment procedure;

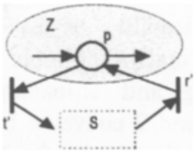


Figure 2. Recovering scheme for the input conditioning method.

- the alternated path method: where an abnormal state detected in place p_i is treated by generating an alternative path from p_i directly to p_m through the sub-net S, in which there are states that recover the system and avoid the execution of activity Q (figure 3). This method is equivalent to the use of redundant resources in the system that substitute the function of Q;

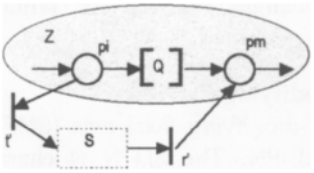


Figure 3. Recovering scheme for the alternated path method.

- the backward recovery method: where, after executed the activity Q, if a failure state is detected in p_m, the system must re-execute Q, i.e., the execution of the operations of the sub-net S imply in the re-execution of Q (figure 4);

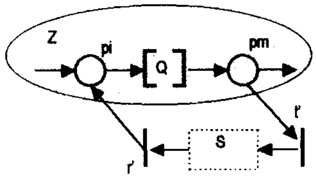


Figure 4. Recovering scheme for the backward recovery method.

- the forward recovery method: where if a failure is detected directly in the activity Q, the system returns a normal state by executing the activities in the sub-net S as shown in figure 5.

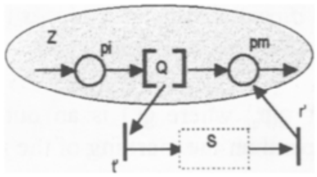


Figure 5. Recovering scheme for the forward recovery method.

These recovering methods augment the number of nodes and connecting arcs in the original net generating more conflict situations. Then, it is commonly assumed that processes with failures are too complex to be included in PN representation.

In fact, the modeling of the normal behavior of the system associated with the failure treatment procedures is impossible in practice when considering the different levels of the details involved in the control system representation. The failure treatment involves the structural changes in the evolution of the system and in the case of PN, a graph re-structure.

In the following, an extension of PN that models structural changes required for failure treatment modeling is presented.

3.2 Self-Modifying Petri Net

The self-modifying Petri nets (SMPN) [9] is an extension of PN. The SMPN is characterized by changes in their structure as function of the current state of system. It has *arcs* with weight k which is a function of the number of *tokens* in specific *places*.

Explaining in a detailed way, the *arc* is labeled with the name of the *place* p_q whose number of *tokens* is associated $k=M(p_q)$.

Then, during the dynamic evolution of a SMPN, represented by the *token* distribution (marking) in the *places*, the weight of arcs can be changed due to changes in the marking. This procedure changes the weight of the *arcs* (temporally or permanently), adding the possibility of assuming the value zero, which means the inexistence of an *arc*.

This feature of SMPN can be employed for structural changes in the normal procedure of the system such as the case of failure and its treatment.

Formally a SMPN can be defined as a quintuple $SMPN = \{P,T,I,O,M_0\}$, where:

- $P = \{p_1, p_2, \dots, p_m\}$ is a finite set of *places*.
- $T = \{t_1, t_2, \dots, t_n\}$ is a finite set of *transitions*.
- $I(p_j,t_i) = f(p_q)$ where $f(.)$ is an input function that defines *directed arcs* from *places* to *transitions* and depends on the marking of *place* p_q, i.e., $f(p_q) \in \mathbf{N}$.
- $O(t_i,p_j) = g(p_u)$ where $g(.)$ is an output function that depend on the marking of the *place* p_u, i.e., $g(p_u) \in \mathbf{N}$.
- M_0 is the initial marking.

This definition includes expressions such as $I(p_j,t_i)=2+M(p_q)^3$ for the weigth of *arcs*. However, in this paper and in the presented references, there are only cases in which $I(p_j,t_i)=M(p_q)$, i.e., in the graphical representations, the *self-modifying arcs* are labeled only with the name of *place* p_q, as shown in fig. 6.

In a SMPN the "firing rule" for a *transition* is, essentially the same as the firing rule for PN. A *transition* ($t_i \in T$) is said to be enabled if each input *places* p_j of t_i (where $I(p_j,t_i)>0$) is marked with at least the same number of *tokens* of the weight of the *arc* $I(p_j,t_i)$. Basically, a *transition* $t_i \in T$ is enabled for firing if and only if:

$M(p_j) \geq I(p_j,t_i) = \; k_{ji}$ if $I(p_j,t_i)$ is a normal *arc*
or $= \; f(p_q)$ if $I(p_j,t_i)$ is a *self-modifying arc*

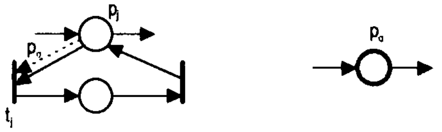


Figure 6. PN with self-modifying arc.

In [9] a SMPN matrix and how the analytical analysis of a system can be conducted is presented. The same work presents the properties of SMPN such as reachability, boundedness and liveness.

Based on the definition of SMPN, there is the possibility of a transition being fired undeterministically if the weight of the input *self modifying arcs* are zero. This situation normally is not proper for manufacturing systems and therefore the following definitions are considered:

- Definition 3.1: a SMPN is said spontaneous, if for each reachable marking, there is a *transition* t_i which is firable and which moves at least one *token* from the place p_j. The word "spontaneous" defines the liveness type (L4-live [7]).
- Definition 3.2: if a case exists, where *transitions* in a SMPN are disabled, but in the equivalent PN (without *self-modifying arcs*) they are enabled, then, this SMPN is a "spontaneous process system with constraints" (SPSC). Therefore, *self-modifying arcs* are added to increase the number of contraints that should be satisfied for firing the *transitions* and thus eliminating any "conflict" [7] it is involved in.

4. APPLICATION EXAMPLE

A example of a manufacturing system is used to illustrate the aplicability of SMPN. The system considered has 2 machining-tools (M1 and M2), 1 robot for loading and unloading parts of the machining-tools, and 2 conveyor belts. The system scheme is shown in figure 8. The system can process 3 types of parts (w1, w2, w3). The w1 parts are processed in M1, the w2 parts are first processed in M1 and after that in M2, the w3 parts are processed only in M2.

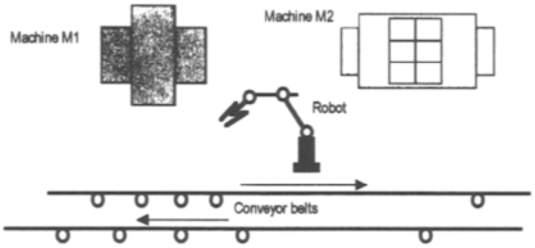

Machine M1
Machine M2
Robot
Conveyor belts

Figure 7. Manufacturing system with 2 machine-tools, 1 robot and 2 conveyor belts.

First, the failures for M1 or M2 are considered. A predictive maintenance procedure is also considered when M1 or M2 are not processing parts.

The SMPN model (type SPSC) of the manufacturing system for processing the parts is shown in figure 8. In the right side of the figure is a model (sub-nets) of the monitored (or diagnosed) states in M1 and M2. The sub-net marking defines the weight of *self-modifying arcs* of the net (in the left side) -- these *arcs* solve the “conflicts” in the SMPN.

The load and unload operations are decomposed in a more detailed set of activities to describe additional failure treatment process. Figure 9 shows a SMPN model (type SPSC) of the machine load/unload operation. This model consider failures in the grasp/move/release operations of the robot and failures by breaking of tools of the machine. The failures considered in the operations are:

- Incorrect positioning of part which not allows a correct grasp.
- Sliding of the part by incorrect force application on the griper.
- Unexpected crash between part/griper and any obstacle which can produce sliding (partial or total) of the part.

The normal states (without failures) are modeled by the horizontal sequence of *places*, this sequence has 3 stages and start when the robot grasps the parts for loading them and finish when the robot unload the part after its processing. In each *place* is considered a failure, thus for each normal state is considered an abnormal state.

The right side of figure 9 show a model of monitored (or diagnosed) states for the machine-tool and for the robot. The marking of these *places* define the weight of *self-modifying arcs* of the main net (in the left side).

5. CONCLUSION

The application of SMPN for modeling of manufacturing system considering failure detection and treatment has been presented. The developed studies [16,17,18] confirm the following points:

- the utilization of SMPN for specifying the control system considering failure treatment is relatively natural, and avoid the explosion of the number nodes in the graph;
- the utilization of *self-modifying arcs* reduce the “conflict” situations;
- the modeling and analysis of complex systems are relatively simpler because SMPN manipulate separated nets (graphically) but connected procedurally;
- the utilization of the same formalism for both the production process modeling and the state monitoring and diagnosis is effective for the analysis of system integration.

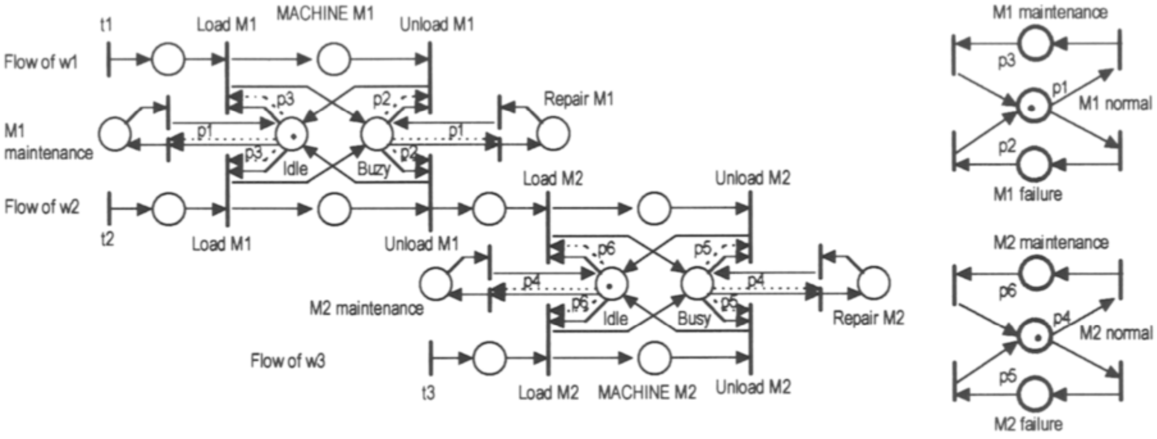

t1
Load M1
MACHINE M1
Unload M1
Flow of w1
M1 maintenance
p3
p2
Repair M1
p1
p1
p3
Idle
Buzy
p2
Flow of w2
t2
Load M1
Unload M1
Load M2
Unload M2
p6
p5
p4
p4
p5
p6
Idle
Busy
M2 maintenance
Repair M2
Flow of w3
t3
Load M2
MACHINE M2
Unload M2
M1 maintenance
p3
p1
M1 normal
p2
M1 failure
M2 maintenance
p6
p4
M2 normal
p5
M2 failure

Figure 8. SMPN of the production process.

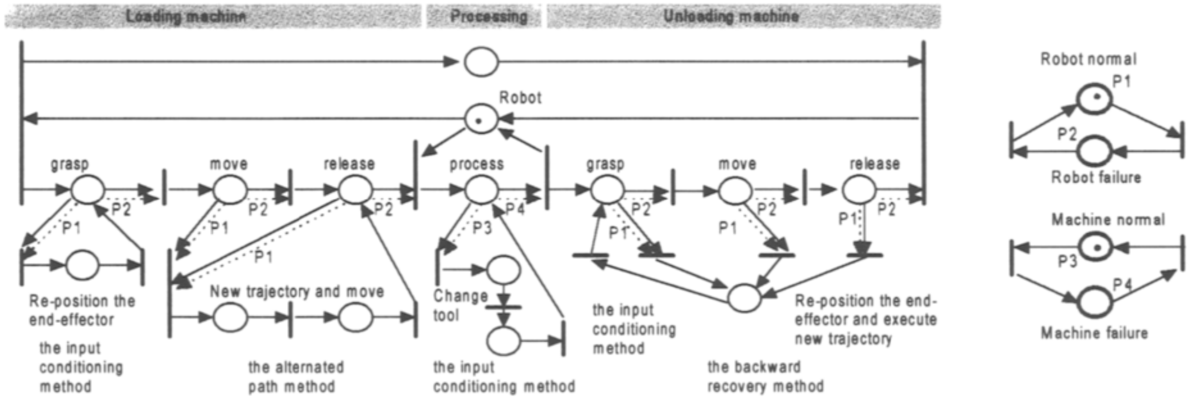

Loading machine
Processing
Unloading machine
Robot
grasp
move
release
process
grasp
move
release
P1
P2
P3
P4
Re-position the end-effector
New trajectory and move
Change tool
the input conditioning method
the alternated path method
the input conditioning method
the input conditioning method
Re-position the end-effector and execute new trajectory
the backward recovery method
Robot normal
Robot failure
Machine normal
Machine failure

Figure 9. SMPN of the loading/unloading operation.

Therefore it has been confirmed that SMPN is adequate for modeling dynamic systems with structural changes.

The authors intend to continue the study of the application of PN and its extensions for this class of systems considering the problem of system design with automatic recovery from errors and failures.

REFERENCES

1. W.Wobbe, Anthropocentric Production Systems are Socio-Technological Innovations, Proc. of IEEE/IFIP BASYS'96, Chapman & Hall, London (1996).
2. I.Kovács and A.Moniz, Issues on the Anthropocentric Production Systems, Proc. of IEEE/IFIP BASYS'95, Chapman & Hall, London 1995.
3. L.M.Camarinha-Matos and H.Afsarmanesh, Towards Balanced Automation, Proc. of IEEE/ IFIP BASYS'95,Chapman&Hall,London 1995.
4. C.Cassandras and P.Ramadge, Toward a Control Theory for Discrete Event Systems, IEEE Control Syst. Magazine, 10 (1990) 4.
5. P.E.Miyagi, Programmable Control: Control of Discrete Event Systems, Edgard Blücher Ed., São Paulo, SP, 1996. (In Portuguese)
6. M.Zhou and F.Dicesare, Petri Net Synthesis for Discrete Event Control of Manufacturing Systems, Kluwer A. Pub., 1993.
7. T.Murata, Petri Nets: Properties, Analysis and Applications, Proc. of the IEEE, 77 (1989) 4.
8. R.Zurawski and M.Zhou, Petri Nets and Industrial Applications: A Tutorial, IEEE Trans. on Industrial Electronics, 41 (1994) 6.
9. R.Valk, Self-Modifying Nets, A Natural Extension of Petri Nets, Advances in Petri Nets, Springer Verlag, Berlin, 62 (1978).
10. U.Rembold, et. al., Computer Integrated Manufacturing and Engineering, Addison-Wesley, 1994.
11. P.Frank, Principles of Model-Based Fault Detection. Proc. of the Int. Symp. on AI in Real-Time Control, Delft (1992).
12. N.Maruyama, Fault Detection in Uncertain Systems Using Neuro-Fuzzy Models, D.Phil. Thesis, Univ. of Oxford, Oxford, 1997.
13. F.V. Jensen, An Introduction to Bayesian Networks. Springer Verlag, New York, 1996.
14. F.Cozman, Robustness Analysis of Bayesian Networks with Local Convex Sets of Distributions. 13th Uncertainty in AI Conference, Providence (1997).
15. Y.Umeda, et. al., A Design Methodology for Self-Maintenance Machines, Journal of Mechanical Design, 117 (1995).
16. M.Sasaki, Application of Self-Modifying Petri Nets in Manufacturing Systems, Master Dissertation, Univ. of São Paulo, São Paulo, SP, 1996. (In Portuguese)
17. L.A.M.Riascos, Modeling of Failure Treatment in Manufacturing Systems through Petri Nets, Master Dissertation, Univ. of Sao Paulo, São Paulo, SP, 1998. (In Portuguese)
18. P.E.Miyagi, et al., Self Modifying Petri Nets for Manufacturing Systems Considering Fails and Maintenance Procedures, 14th Brazilian Congress of Mechanical Engineering, Bauru, (1997).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Synthesis and analysis of feedback linearization for pneumatic actuator control systems

Fulin Xiang and Jan Wikander

Department of Machine Design, Royal Institute of Technology
SE - 100 44 Stockholm, Sweden
Email: xiang@damek.kth.se, jan@damek.kth.se

Abstract In this paper a robust feedback linearization methodology is introduced for a pneumatic actuator servo control systems. By the presented method, an effective friction compensation, with the LuGre dynamic friction mod el, was obtained even though there is considerable dynamics between the input control signal and the manipulating force output. High motion and force control performance was achieved with the presented techniques. Stability and robustness analysis with respect to uncertainties and linearization residual are included. Experimental motion and force control results are also presented.

1. INTRODUCTION

Even though many control approaches have been suggested for pneumatic servo actuators, it is still very difficult to achieve high control performance due to the substantial inherent nonlinearities. The nonlinearities within a pneumatic actuator mainly include nonlinear friction, air flow and pressure build-up nonlinearities, coupling between pressure build-up and piston motion, dead-zone and saturation.

To cope with these nonlinearities, a nonlinear control methodology has to be applied. In recent years there have been some significant advances in the area of nonlinear control system design, such as feedback linearization [1-2] and backstepping [3]. In practice, the big obstacle for applying these nonlinear design techniques is the requirement that the nonlinearities must be continuously differentiable a sufficient number of times [4]. In the pneumatic servo control case, the nonlinear friction exhibits, in the vicinity of zero velocity, fast dynamics, which, in combination with limited sampling rate makes differentiation impossible. Position and pressure measurements are also not suitable for differentiation more than two times.

In [5] the feedback linearization technique was applied in a pneumatic servo by means of separate inner pressure control loop and outer motion control loop, resulting in a good control performance. A nonlinear modelling and identification method was introduced in [6] for the purpose of reducing the feedback linearization residual and improve control quality.

This paper extends the work in [5, 6] with analysis of stability and robustness with respect to linearization residual and other uncertainties. Further, a friction compensation, with the LuGre dynamic friction model [7], is shown to be effective despite considerable dynamics between friction input signal and the force controller output. Stabilization is derived with the existence of linearization residual both for the inner pressure loop and outer motion control loop, with which robustness can be achieved for the linearized closed-loop system. Experimental motion and force control results are also presented.

2. MODELLING AND FEEDBACK LINEARIZATION

With the assumptions that the supply pressure and temperature are known constants, the state transitions in the chambers constitute polytropic processes with constant polytropic exponent, air can be considered as an ideal gas, the kinetic energy of the air is negligible and, finally, a horizontally mounted rodless cylinder with two servo valves is considered, the pneumatic actuator system can be described by the following equations.

- Motion equation

$$m\ddot{y} = A_c(p_1 - p_2) - F_{fr} = F - F_{fr} \qquad (1)$$

where m is the moving mass, A_c the cylinder piston cross area, p_1 and p_2 the two chamber pressures, y cylinder piston position and F_{fr} the friction force.

- LuGre friction force mode

$$F_{fr} = \sigma_0 z + \sigma_1 \frac{dz}{dt} + F_v v$$
$$\frac{dz}{dt} = v - \frac{|v|}{g(v)} z \qquad (2)$$

$$g(v) = \frac{1}{\sigma_0}\left(F_c + (F_s - F_c)e^{-(v/v_s)^2}\right)$$

where $v = \dot{y}$ is the velocity, z virtual friction state, σ_0 stiffness coefficient, σ_1 internal damping coefficient, F_v viscous friction, F_c Coulomb friction, F_s static friction and v_s Stribeck velocity.

- Pressure build-up equations

$$\dot{p}_1 = \left(\frac{nRT}{A_c}Q_1 - nvp_1\right)\frac{1}{y + y_0} \tag{3}$$

$$\dot{p}_2 = \left(\frac{nRT}{A_c}Q_2 + nvp_2\right)\frac{1}{l - y + y_0} \tag{4}$$

where R is gas constant, n polytropic exponent, T temperature of the pressure supply, l cylinder stroke and Q air flow rate and y_0 the effective initial position corresponding to the non-working volume.

- Flow rate equations

$$Q = uk_v(u)G_v(D)p_u\sqrt{\frac{1-b}{RT}}f(p_r, b) \tag{5}$$

$$f(p_r, b) = \begin{cases} 1 & p_r = \dfrac{p_d}{p_u} \le b \\ \sqrt{1 - \left(\dfrac{p_r - b}{1 - b}\right)^2} & p_r = \dfrac{p_d}{p_u} > b \end{cases}$$

where u is the valve input, $k_v(u)$ static characteristics of the valve including the static nonlinearities (dead-zone, saturation, etc.), $G_v(D)$ normalized valve dynamics, b critical pressure ratio, p_u and p_d up and down stream pressure, D Laplace operator s or differential operator p. Let

$$L_p = nRT/A_c \tag{6}$$

$$L_Q(D) = \begin{cases} G_v(D)p_u\sqrt{(1-b)/(RT)}, \text{ charging} \\ G_v(D)\sqrt{(1-b)/(RT)}, \text{ discharging} \end{cases} \tag{7}$$

$$N_{v1} = -nvp_1 \qquad N_{v2} = nvp_2 \tag{8}$$

$$N_{y1} = 1/(y + y_0)$$
$$N_{y2} = 1/(l - y + y_0) \tag{9}$$

$$N_Q = \begin{cases} k_v(u)f(p_r, b), \text{ charging} \\ p_u k_v(u)f(p_r, b), \text{ discharging} \end{cases} \tag{10}$$

where L_p, L_Q represent the linear parts and N_y, N_v and N_Q represent the nonlinear parts. Using the corresponding coefficients 1, 2 for the respective chambers the pressure build-up model can be expressed as

$$\dot{p} = (uL_pL_Q(D)N_Q + N_v)N_y \tag{11}$$

Assuming N_Q, N_y and L_p are invertible and $L_Q(D)$ is a minimum phase system, u can be chosen as

$$u = \upsilon_p\hat{N}_Q^{-1}\hat{N}_y^{-1} - \hat{N}_v\hat{L}_p^{-1}\hat{L}_Q^{-1}\hat{N}_Q^{-1} \tag{12}$$

Substituting this equation into Eq. 11, the resulting pressure build-up dynamics is linearized as

$$\dot{p} = \upsilon_p\tilde{N}_Q\tilde{N}_yL_pL_Q - \hat{N}_v\tilde{L}_p\tilde{L}_Q\tilde{N}_QN_y + N_vN_y \tag{13}$$

where $\tilde{L}_p = L_p\hat{L}_p^{-1}$, $\tilde{L}_Q = L_Q\hat{L}_Q^{-1}$, $\tilde{N}_Q = N_Q\hat{N}_Q^{-1}$,

$\tilde{N}_y = N_y\hat{N}_y^{-1}$ the linearization residual and υ_p is the equivalent control signal for the linearized system. When the items, labelled with the over hat (^), derived from estimated or measured parameters tend to their corresponding true values, the inverse for the linear dynamics and nonlinearities shown in Equation 12 can be exactly realized, then the linearized pressure build-up system becomes

$$\dot{p} = \upsilon_pL_pL_Q(D) \tag{14}$$

Based on this linearized pressure build-up model, an inner force or pressure controller can be designed by traditional linear feedback control design techniques. Let $G_F(D)$ represent the closed force control loop including both chambers, Then the force response can be expressed as

$$F = F_{ref}G_F(D) \tag{15}$$

Assuming $G_F(D)$ is a minimum phase system, the force reference can be chosen as

$$F_{ref} = \upsilon_m + \hat{F}_{fr}\hat{G}_F^{-1}(D) \tag{16}$$

where υ_m is the equivalent control signal for the linearized system. Substituting this equation into Eq. 1 gives the linearized motion control system as

$$m\ddot{y} = \upsilon_mG_F(D) - \tilde{F}_{fr} \tag{17}$$

where

$$\tilde{F}_{fr} = F_{fr} - \hat{F}_{fr}G_F(D)\hat{G}_F^{-1}(D) \tag{18}$$

represents the friction compensation residual. When the friction estimate tends to its true value this residual tends to zero, which gives

$$m\ddot{y} = \upsilon_mG_F(D) \tag{19}$$

3. LINEARIZATION RESIDUAL AND STABILIZATION

In practice it is impossible to get rid of uncertainty. What we can do is only to reduce the uncertainty. Large uncertainty may cause large linearization residual, and the residual may also be due to the implementation capacity. For example, the inverse of some dynamics may not be possible to be implemented due to implementation constraints, potentially resulting in large residual. The linearization residual acts as a perturbation to the linearized system. Even a small value of this kind of perturbation may be harmful to the linearized system and even induce instability. To illustrate this let us analyse some simplified cases and see what kind of measures can be taken to guarantee robustness.

3.1 Inner pressure control loop

Considering chamber 1 and $v > 0$ case, when neglecting the uncertainty and dynamics in the first term of the right hand side of Eq. 13 and assuming that there are no uncertainties in the measured variables y, p and the derived variable v, Eq. 13 is rewritten as

$$\begin{aligned} \dot{p}_1 &= \theta_{p1}\upsilon_{p1} - np_1 v(1 - \tilde{L}_p\tilde{L}_Q\tilde{N}_Q)N_y \\ &= \theta_{p1}\upsilon_{p1} - p_1 v\Delta_{p1}(\upsilon_{p1}, y) \end{aligned} \tag{20}$$

where $\theta_p > 0$ is a constant gain and $\Delta_{p1}(\upsilon_{p1}, y)$ the bounded uncertainty due to the linearization residual. The uncertainty is mainly considered as a function of the input signal υ_{p1} and position $y>0$. As the residual tends to zero the uncertainty will also tend to zero. Selecting $\upsilon_{p1} = \theta_{p1}^{-1}\dot{p}_{d1} - k\tilde{p}$ where p_{d1} is the pressure reference, $\tilde{p}_1 = p_1 - p_{d1}$ is the control error and $k>0$ the feedback gain, giving

$$\dot{\tilde{p}}_1 + k\theta_{p1}\tilde{p}_1 = -p_1 v\Delta_{p1}(\upsilon_{p1}, y) \tag{21}$$

Introducing deviation variables $\delta p_1 = p_1 - p_1^o$, $\delta\tilde{p}_1 = \tilde{p}_1 - \tilde{p}_1^o$, and $\delta p_{d1} = p_{d1} - p_{d1}^o$ where the superscript ^{o} denotes the steady-sate operating points. Assuming the pressure reference p_{d1} is a constant then $p_{d1} = p_{d1}^o$ and $\delta p_1 = \delta\tilde{p}_1$, Eq. 21 can be written as

$$\delta\dot{\tilde{p}}_1 + k\theta_{p1}\delta\tilde{p}_1 = -\delta\tilde{p}_1 v\Delta_{p1}(\upsilon_{p1}, y)$$

Replacing $\delta\tilde{p}_1$ by $\tilde{p}_1$ and after rearranging we have

$$\dot{\tilde{p}}_1 + [k\theta_{p1} + v\Delta_{p1}(\upsilon_{p1}, y)]\tilde{p}_1 = 0 \tag{22}$$

For under compensation, by which we mean that compensation is less than the influence of the corresponding nonlinearities, *i.e.* $1 - \tilde{L}_p\tilde{L}_Q\tilde{N}_Q \geq 0$ or $\Delta_{p1}(\upsilon_{p1}, y) \geq 0$, the closed-loop system, Eq. 22, is stable. For over compensation, $\Delta_{p1}(\upsilon_{p1}, y) < 0$. In this case if v is independent of $\tilde{p}_1$, while $k\theta_{p1} \geq v|\Delta_{p1}(\upsilon_{p1}, y)|$, the system is still stable. But v is directly related to $\tilde{p}_1$ through an integration. So the real situation is when over compensation of N_v, $\Delta_{p1}(y, \upsilon_{p1}, d_1) < 0$, the perturbation part is counteracting the negative feedback part $k\theta_{p1}$, this may cause an increase of $\tilde{p}_1$, the increase of $\tilde{p}_1$ may induce the increase of v, the increase of v increases the counteraction of the negative feedback again, potentially leading to instability. So over compensation of N_v may induce instability. This is especially the case when there is no outer motion control loop. This conclusion is valid for both chambers and both motion directions, and can be proven similarly.

From Eq. 12 it can be seen that if we use the upper bound for the inverse compensator estimation and the lower bound for the non-inverse compensator estimation, then the over compensation can be avoided. In this case Eq. 12 can be rewritten as

$$u = \upsilon_p\hat{N}_Q^{-1}\hat{N}_y^{-1} - (\hat{N}_v)_l(\hat{L}_p)_u^{-1}(\hat{L}_Q)_u^{-1}(\hat{N}_Q)_u^{-1} \tag{23}$$

where the subscript $_l$ indicates the lower bound and $_u$ indicates the upper bound.

3.2 Outer motion control loop

The LuGre friction model can explicitly capture the fast dynamics in the transition from sticktion to sliding. The model normally requires a high sampling rate for the sake of stability. In practice, due to the limitation of sampling rate, the differentiation of the friction estimation, normally can not be implemented. The inverse of $G_F(D)$ in Eq. 16 at least leads to the need for one differentiation of the friction estimation. So the inverse of $G_F(D)$ has to be omitted. Supposing that the estimation of the friction is equal to the true value, Eq. 18 becomes

$$\tilde{F}_{fr} = F_{fr}(1 - G_F(D)) \tag{24}$$

Since $G_F(D)$ is a stabilized strictly proper system we can approximate the friction compensation residual in the following form

$$\tilde{F}_{fr} = \hat{F}_{fr}w_{fr}(s)\Delta(s) = \hat{F}_{fr}\frac{M_h s + M_l\omega_E}{s + \omega_E}\Delta(s) \tag{25}$$

where $w_{fr}(s)$ is a weighting function and $\|\Delta\|_\infty \leq 1$ denotes the normalized perturbation, in which the uncertainty is involved as depicted in Figure1. Referring to Eq. 18 the notion of over compensation means that the compensation is larger than the influence of the non-linearity, i.e. $\tilde{F}_{fr} < 0$. So for the worst over compensation $\Delta = -1$, and for the worst under compensation $\Delta = 1$. For simplicity, we neglect $G_F(D)$ in Eq. 17 and choose the control signal as

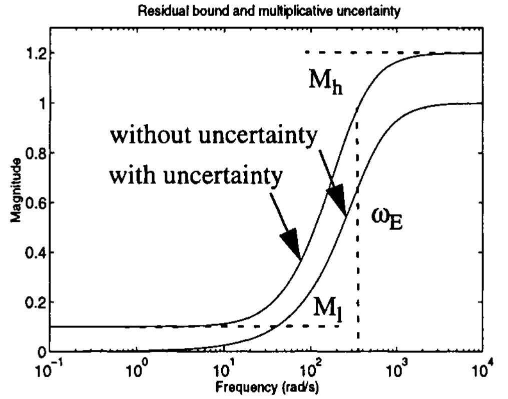


Figure 1: Residual and multiplicative uncertainty

$$\upsilon_m = m\ddot{y}_d - k_1\dot{\tilde{y}} - k_0\tilde{y} \tag{26}$$

where y_d is the position reference, $k_1>0$, $k_2>0$ are the feedback control gains and $\tilde{y} = y - y_d$ is the position control error. Here we also suppose that the y_d at least can be differentiated two times which is ensured by a suitable trajectory planner. Then, the closed loop motion control system can be expressed as

$$m\ddot{\tilde{y}} + k_1\dot{\tilde{y}} + k_0\tilde{y} = \tilde{F}_{fr} = \pm\hat{F}_{fr}w_{fr}(s) \tag{27}$$

Let us focus our discussion for operation around a steady-state point where $v^o=v_d=0$, $y^o=y_d$. In this region, due to the fast friction dynamics, the most substantial friction compensation residual is supposed to appear. Assuming a small $\tilde{y}$ which thus also can represent the friction state z, the friction compensation residual can be expressed as

$$\tilde{F}_{fr} = \pm[\sigma_0\tilde{y} + f_v(v)\dot{\tilde{y}}]w_{fr}(s) \tag{28}$$

where f_v represents the viscous friction coefficient. In the Stribeck effect region $f_v < 0$. Substituting Eq. 28 into Eq. 27, and rearranging we have

$$m\tilde{y}^{(3)} + K_2\tilde{y}^{(2)} + K_1\tilde{y}^{(1)} + K_0\tilde{y}^{(0)} = 0 \tag{29}$$

where K_0, K_1 and K_2 are shown in Eq. 30 - 32. Since $m > 0$, according to Routh-Hurwitz criterion, the necessary and sufficient condition for all the roots of Eq. 29 lie in the left half of the s-plane is that all the elements of the first column of the Routh tabulation are of positive sign, *i.e.*

$$K_0 = (k_0 \mp \sigma_0 M_l)\omega_E > 0 \tag{30}$$

$$K_2 = k_1 + m\omega_E \mp f_v(v)M_h > 0 \tag{31}$$

$$K_1 - mK_0/K_2 = k_0 + k_1\omega_E \mp \sigma_0 M_h \mp f_v(v)M_l\omega_E$$
$$-\frac{m\omega_E(k_0 \mp \sigma_0 M_l)}{k_1 + m\omega_E \mp f_v(v)M_h} > 0 \tag{32}$$

To give an intuitive feeling of how the linearization residual and uncertainty affect the stability of the closed-loop system, let's check some specific situations.

i. When there is no uncertainty and linearization residual, we have $M_l=0$, $M_h=0$, $\omega_E \to \infty$. Obviously all the inequalities 30 to 32 are satisfied.

ii. Under compensation means applying the positive sign of $\mp$ term in Eq. 30 to 32. For the the non-Stribeck effect region $f_v(v)\geq 0$ and thus Eq. 30 and 31 are satisfied. The last term in the right hand of Eq. 32 has the following inequality,

$$\frac{m\omega_E(k_0 + \sigma_0 M_l)}{k_1 + m\omega_E + f_v(v)M_h} < \frac{m\omega_E(k_0 + \sigma_0 M_l)}{m\omega_E}$$
$$= k_0 + \sigma_0 M_l \tag{33}$$

noting also that $M_h \geq M_l$ and comparing Eq. 32 with 33, it can be seen that the inequality 32 is satisfied. So, in the under friction compensation case, the system is stable for all non-Stribeck effect region.

iii. In the Stribeck effect region $f_v(v) < 0$, for under compensation situation, there will be a limitation on the feedback gain. From Eq. 31 and 32, it can be seen that the increase of k_1 is in the direction of stabilizing the system. For example, choosing $k_1 \geq |f_v(v)|M_h$ can guarantee the system stability.

iv. For the stability in the over compensation situation, both feedback gains k_1 and k_2 have to be larger than certain values. For example, choosing $k_0 \geq \sigma_0 M_l$ Eq. 30 can be satisfied. In the non-Stribeck effect region, selecting $k_1 \geq |f_v(v)|M_h$ and noting the inequality 33, we have the inequality 34 or 35.

$$(k_1 - f_v M_l)\omega_E - \sigma_0(M_h - M_l) > 0 \tag{34}$$

$$
\begin{aligned}
& f_v \omega_E (M_h - M_l) - \sigma_0 (M_h - M_l) \\
& = (f_v \omega_E - \sigma_0)(M_h - M_l) > 0
\end{aligned}
\tag{35}
$$

It is clear that when $f_v(v)\omega_E > \sigma_0$, the inequalities can be satisfied. If $f_v(v)\omega_E < \sigma_0$ form Eq. 34 it can be seen that there will be a further requirement on increasing k_1, which shows that the increase of ω_E will contribute to stabilizing the closed-loop system.The Stribeck effect region is clearly less severe due to the positive sign selection in Eq. 30 and 32.

v. When the Routh-Hurwitz stability condition is satisfied in the over friction compensation situation, it can be proven that it will also be satisfied in the under friction compensation situation.

3.3 Asymmetric pressure control strategy

In the above analysis two requirements are derived for stabilization of the friction compensated closed-loop system. One is that the feedback gains k_1 and k_2 should be larger than certain values. With a certain given positioning reference this means that $|\upsilon_m|$ or the $|F_{ref}|$ can be larger than certain value without conflicting with the input constrain. Another requirement is that the dynamics (or band width) of $G_F(D)$ should be as large as possible so that we can have $M_l \rightarrow 0$, $\omega_E \rightarrow \infty$. The limitation of the band width of $G_F(D)$ is mainly due to the constraint of the input signal, *i.e.*

$$
|u| = \left|\upsilon_p \hat{N}_Q^{-1} \hat{N}_y^{-1} - \hat{N}_v \hat{L}_p^{-1} \hat{L}_Q^{-1} \hat{N}_Q^{-1}\right| \le M_{u0} \tag{36}
$$

which also can be written as

$$
\left|\upsilon_p \hat{N}_Q^{-1} \hat{N}_y^{-1}\right| \le M_{u0} - \left|\hat{N}_v \hat{L}_p^{-1} \hat{L}_Q^{-1} \hat{N}_Q^{-1}\right| \le M_{u1} \tag{37}
$$

As a result of the feedback linearization of Eq. 12 we get the two chamber pressure build-up dynamics of Eq 14 independent of the piston motion, but meanwhile we also have a position varying constraint on υ_p as shown by Eq. 38.

$$
|\upsilon_p| \le \hat{N}_Q \hat{N}_y M_{u1} \le \hat{N}_y M_u = M_\upsilon \tag{38}
$$

According to Eq.38 for chamber 1, M_υ increases with the decrease of position, with the opposite situation in chamber 2. Representing the linearized pressure build-up of Eq. 14 by transfer function P and the cascade feedback controller with K, we have $\upsilon_p = Kp_{ref}(1 + KP)^{-1}$. This gives that an increase of M_υ gives the possibility either to increase K, which means increase of the bandwidth of $G_F(D)$, or to increase p_{ref}. In order to fully use of the system resources, an asymmetric pressure control strategy was adopted, in which the pressure reference was chosen as

$$
\begin{aligned}
p_{1ref} &= p_0 + \left(1 + k_{pr}\frac{l-y}{l}\right)\frac{F_{ref}}{2A_c} \\
p_{2ref} &= p_0 - \left(1 + k_{pr}\frac{y}{l}\right)\frac{F_{ref}}{2A_c}
\end{aligned}
\tag{39}
$$

where $k_{pr} \ge 0$ is a constant gain and p_0 the base pressure. A more detailed discussion about the selection of p_0 is given in [5].

4. EXPERIMENTAL RESULTS

The pneumatic actuator system used for the experimental test contains a 32 × 400 rodless cylinder, two servo valves, two pressure sensors and an optical incremental position sensor with a resolution of 5 μm. The maximum static friction for $p_0 = 4e5\ Pa$ is about 15% of the maximum static cylinder force.

A square wave positioning of $y_{ref} = 0.2 \pm 0.1\ m$ with $v_{ref} = 1.2\ m/s$ is shown in Figure 2. The upper figure is the comparison of the displacement and its

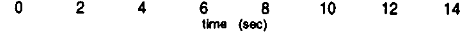


Figure 2: Square wave positioning at $v = 1.2\ m/sec$

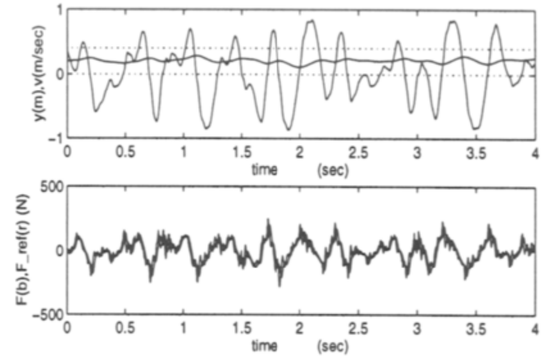


Figure 3: Force control with the varying position and velocity

reference while the lower one shows the positioning error. It can be seen that the steady positioning error goes to zero which means that it is less than the sensor resolution. The force control results are shown in Figure 3, where the upper figure shows the varying operation condition (piston velocity and position) for a randomly chosen force reference signal, the lower is the comparison of force reference and response.

Two extreme positioning cases for the pneumatic servo are low velocity and small distance positioning. In low velocity positioning the dominant friction is in the mode of pre-sliding and sticktion. Large model uncertainties and compensation error may occur in this situation. Despite of this, the steady positioning error can also go below the 5 μm position sensor resolution as shown in Figure 4. Small positioning means small displacement, small pressure reference and thus small control input. This means that the uncertainty will be substantial. In this case, the nonlinearities and parameter variations, such as dead zone or the nonlinearities of $k_v(u)$, leakage and zero point drift etc. may dominate the system characteristics and make the small positioning difficult. The better the nonlinearities compensation, the smaller step positioning can be realized for a certain pneumatic system. Figure 5 shows a positioning of $y_{ref} = 0.35 \pm 0.0005$ *m*. The steady positioning error can also easily be kept below *5 μm*.

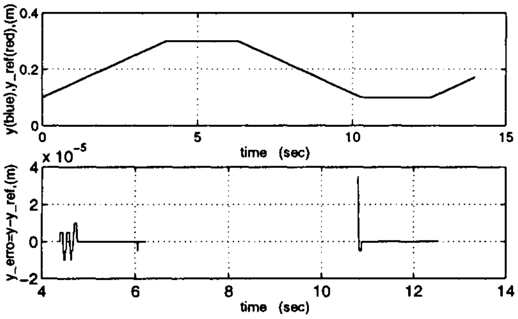


Figure 4: Low velocity positioning

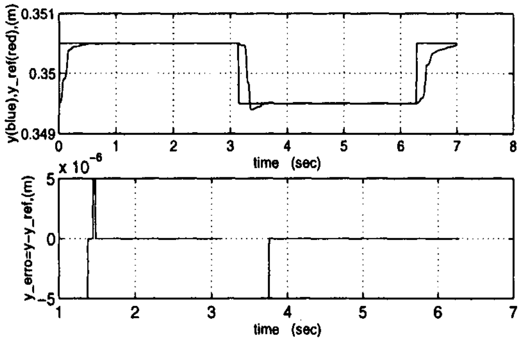


Figure 5: Small square wave positioning

5. CONCLUSION

High pneumatic actuator motion and force control performance can be achieved with the presented feedback linearization techniques, which were implemented separately in an inner force control loop and an outer motion control loop. An effective friction compensation, with the LuGre dynamic friction model, was obtained, although there is considerable dynamics between the control input signal and the force output. Linearization residual, which are mainly due to the estimation uncertainties and the implementation constraints, can induce system instability. Stabilizing techniques were introduced for both the inner force and outer motion control loop, with which robustness can be achieved for the linearized closed-loop system.

REFERENCES

1. Isidori, A., Nonlinear control systems, 3rd ed., Springer-Verlag, Berlin, 1995
2. Slotine, J., W. Li, Applied Nonlinear Control, Prentice Hall, Englewood Cliffs, 1991
3. Krstic, M., I. Kanellakopoulos, P. V. Kokotovic, Nonlinear and adaptive control design, John Wiley & Sons, New York, 1995
4. Tao, G., P. V. Kokotovic, Adaptive Control of Systems with Actuator and Sensor Nonlinearities, John Wiley & Sons, New York, 1996
5. Xiang, F., J. Wikander, B. Eriksson, Nonlinear control of pneumatic servo - a feedback linearization approach, Proceedings of the 1st International Fluid Power Conference, 1998, Aachen, Germany.
6. Xiang, F., J. Wikander, Nonlinear modelling and identification of pneumatic servo, Proceedings of the 17th IASTED International Conference on Modelling, Identification and Control, Grindelwald, Switzerland, 1998
7. Canudas de wit, P. Lischinsky, Adaptive friction compensation with dynamic friction model, Proceedings of the 13th IFAC Triennial World Congress, San Francisco, USA, 1996
8. Kanellakopoulos, I., Robust nonlinear control design with input and measurement disturbance, Proceedings of the 1997 American control conference, Albuquerque, NM, June 1997

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

On line autotuning for fuzzy logic controlled pneumatic positioners

G. Mattiazzo[a], S. Mauro[a], M. Velardocchia[a] and T. Raparelli[b]

[a]Department of Mechanics, Politecnico di Torino
C.so Duca degli Abruzzi 24, 10128 Torino, Italy

[b]Department of Energetics, Università di L'Aquila
Roio Poggio, L'Aquila, Italy

Pneumatic servoaxis driven by PWM piloted digital valves are a suitable economic solution in all those cases in which high precision is not required. Unfortunately this kind of system is strongly non-linear, and hence fuzzy logic is a powerful instrument to solve the control problem. The application of this technique requires the definition of dedicated membership functions for the fuzzy sets, while a unique set of control rules can be defined for every kind of actuator. An algorithm for the automatic tuning of the membership function was developed and tested.

1.INTRODUCTION

The problem of the automatic tuning of fuzzy controllers was deeply investigated by a large number of authors in the last years.

In most cases the large problem of tuning a suitable controller for a namely unknown system was analysed.

A widely adopted technique consists in the application of neural network based learning systems to the fuzzy controller synthesis [1] [2] [3] [4].

Other authors purposed method for direct research of an optimal rule group and of the related membership functions. In a first solution rules are be generated according to a preliminary defined general behaviour of the system by suppression or inhibition of rules where conflicts are found [5]. In this cases the learning algorithm must be driven by the user to choice the dominant behaviour.

In other solutions the rules are extracted analysing the input-output pairs and updated according to the system actual behaviour with the tested rule [6], [7]. Rules and membership functions can be updated just on the basis of the obtained performance, or also considering some characteristic parameters of the controlled system, as, for instance, the open loop static gain.

Beside this work the problem of tuning a fuzzy controller for a pneumatic servoaxis is analysed.

The servoaxis is driven by pulse width modulation piloted valves, and it hence characterised by strong non-linearities.

The problem of controlling this kind of system was deeply investigated by the authors in former works, considering both position control [8], [9], and force control for robotic applications [10], [11].

The acquired experience led to the identification of a general table of rules which must be fitted on the actual system defining each time the membership functions of the input output quantities.

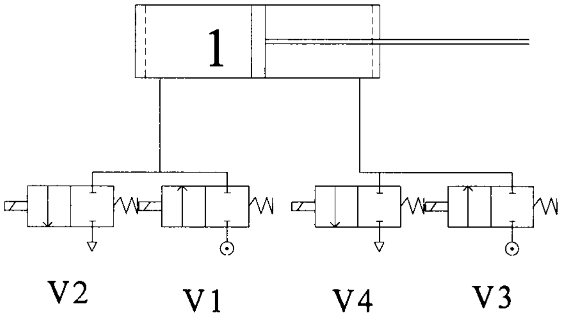
The purposed autotuning algorithm applies hence empirical rules which were previously found and used for trial-and-error manual tuning.

A method for the evaluation of the obtained performance is exposed together with the membership functions update algorithm.

The generated membership functions are stored in the membership function space and updated on line in order to allow continuous running of the servoaxis while tuning. The performance evaluation method and the membership functions update algorithm can run during all the operating life of the servoaxis, allowing hence control adaptability to strong variation of environmental conditions, such as, for instance, seals degeneration.

2. SYSTEM DESCRIPTION

The considered servoaxis (fig. 1) is composed by a pneumatic double effect cylinder (1) supplied and exhausted by two couples of high frequency, two ways, two positions digital valves (V_1,V_2,V_3 and V_4). Finally a position transducer detects the rod displacement. The control loop is closed by a 12 bit fuzzy controller installed on a PC equipped with suitable I/O boards.



The valves are driven in pulse width modulation (PWM), so that their conductance can be considered ideally proportional to the duty cycle. Actually it must be pointed out that because of the valves response time dead bands around the extreme values of the duty cycle must be expected.

The used valves have eight poppets each, which can be convoyed either in one or two ways. A single valve is used to provide supply to both the chamber, while each exhaust is controlled by a dedicated valve. Hence the maximum conductance of each exhaust valve is double than the corresponding supply valve one.

The controller computes the valve duty cycle elaborating the position error and, optionally, its derivative.

Two outputs are provided, a first one for the rear chamber valves and a second one for the front chamber.

Each value is included in the 0 – 4095 range and it is converted in the duty cycle according to the following equations:

$$\frac{t_1}{T}=0 \qquad \frac{t_2}{T}=1-\frac{u}{2048} \qquad if \quad u<2048 \tag{1}$$

$$\frac{t_1}{T}=\frac{u}{2048}-1 \qquad \frac{t_2}{T}=0 \qquad if \quad u\geq 2048 \tag{2}$$

where t1 and t2 the opening times for the supply and exhaust valve respectively, T is the PWM period and u is the output quantity.

The frequency of the PWM can be chosen in a range between 66 and 33 Hz according to the cylinder cut off frequency and to the expected disturbance frequency. In the studied case this parameter was set to 50 Hz. An higher value involve the maximum refresh signal frequency to be higher too, while a lower value let the dead bands relative width to be lower.

The described layout can be applied to cylinders of different stroke and course, and it can be considered a general method for the realisation of pneumatic servoaxis.

The parameter of the used test rig are reported in table 1.

Cylinder course	200 mm
Cylinder diameter	50 mm
Rod diameter	20 mm
Supply valve maximum conductance	$1.19\cdot10^{-8}$ Nm^3/sPa
Exhaust valve maximum conductance	$2.38\cdot10^{-8}$ Nm^3/sPa
PWM period	20 ms
Valve response time (open)	2 ms
Valve response time (close)	1 ms

Table 1 – System parameters

3. CONTROL STRATEGY

The fuzzy controller elaborates two inputs, i. e. position error e and its change in the PWM period Δe.

The position error is computed as

$$e=\frac{(x-x_{rif})K_T K_{AD}}{2}+2048 \tag{3}$$

where x is the actual position, x_{rif} the reference position, K_T the transducer gain and K_{AD} the analogue/digital converter gain. The term 2048 translates the error origin in the middle of the 12 bit range of the fuzzy controller.

The error change in the PWM period is computed as

$$\Delta e = \frac{e_i - e_{i-1}}{T} K_{\Delta e} + 2048 \qquad (4)$$

where e_i is the error at instant i, T is the PWM period and $K_{\Delta e}$ is a suitable gain, which in this application was set to 0.05.

It must be remembered that the control loop must be synchronised with the PWM, and hence its period must be T. The quantity Δe is hence a good approximation for the error derivative.

A rule table to be used with every kind of cylinder was identified (tab 2).

		Δe									
		NM		NS		ZR		PS		PM	
e	NL	PM	NL	PM	NL	PM	NL	PM	NL	PM	NL
	NM	PM	NL	PM	NM	PM	NM	PM	NS	PM	NS
	NS	PM	NS	PM	NS	PM	ZN	PS	ZN	PS	ZN
	ZR	ZP	ZN	ZP	ZN	ZR	ZR	ZN	ZP	ZN	ZP
	PS	ZN	PS	ZN	PS	ZN	PM	NS	PM	NS	PM
	PM	NS	PM	NS	PM	NM	PM	NM	PM	NL	PM
	PL	NL	PM	NL	PM	NL	PM	NL	PM	NL	PM

Table 2 - Rules

The reported rule make the valves to be fully open during the approach run, in order to obtain the maximum available speed, and progressively close them when the rod is close to the reference position.

As the control acts on the valves conductance, the main action is carried out by reducing flow across the exhaust valve of the low pressure chamber, coherently with the pneumatic cylinder speed control technique.

On the other side the action carried out on the high pressure chamber supply valve has a lower effect, as the incoming flow with high chamber pressure and low rod speed is very low.

To apply successfully the above mentioned rules to each kind of cylinder it is necessary to define properly the membership functions of the involved equations. In fact they constitute the actual link between the qualitative description of the controller behaviour and its numerical output.

As the optimal tuning of the membership functions is a long and skilfulness requiring operation, the automatic tuning algorithm object of this work operates on them.

4. MEMBERSHIP FUNCTIONS AUTOTUNING ALGORITHM

The autotuning algorithm operates starting from a standard set of membership functions which forecast the inner 30% of the error range to be described by 5 fuzzy equidistant subset and the remaining 70% by two more subset (NL and PL). According to the previously shown rules this involves the cylinder to reach its maximum speed during the approach run.

The error change is described by 5 fuzzy subset.

Output quantities are described by 7 fuzzy subset, 3 for the supply range and 4 for the exhaust (i.e. value lower than 2048). Their membership functions are singleton, and initially they are disposed in an equidistant way within the active range of supply and exhaust, i.e. excluding the central zone which is supposed to be covered by the dead bands. As dead bands width is not known, it is initially supposed to be extended in a range equal to the 35% of the full output range and centred on the closed valves point.

The cylinder runs for 6 seconds starting from the retracted rod position driven by the described controller, and error and its change are acquired and stored. The data are analysed in order to evaluate the performance of the positioner and to compute corrections for the membership functions.

In a first phase the history of error is analysed to select only points acquired after the approach run, i.e. with position higher than the 90% of the medium value in steady conditions, or higher than the reference if the medium value in steady condition is over the reference but the 90% of its value is not.

The selected data are used to analyse the trajectory in the e Δe plane. The typically obtained result is reported in fig. 2. Usually the first attempt controller set up makes the rod oscillate around a position far from the reference one.

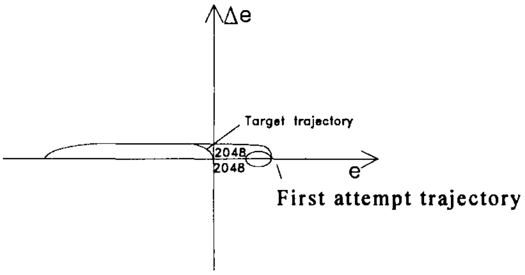


Fig.2 – Typical behaviour of the system at first attempt and target behaviour

When steady conditions are reached the trajectory is enclosed within an ellipse whose semiaxis can be computed as

$$a = \frac{e_{max} - e_{min}}{2} \qquad (5)$$

$$b = \frac{\Delta e_{max} - \Delta e_{min}}{2} \qquad (6)$$

and whose centre C is

$$C = \left(\frac{e_{max} + e_{min}}{2}, \frac{\Delta e_{max} + \Delta e_{min}}{2} \right) \qquad (7)$$

The target trajectory should be characterised by

$a = 0$
$b = 0$
$C = (2048{,}2048)$

The semiaxis length are strictly related between themselves, and the ordinate of point C is always very close to 2048. The obtained performance can hence be evaluated considering an oscillation index I_o and an error index I_e defined as:

$$I_o = \frac{\Delta e_{max} - \Delta e_{min}}{2} \qquad (8)$$

$$I_e = \frac{e_{max} + e_{min}}{2} - 2048 \qquad (9)$$

The membership functions are then corrected reducing the base of the triangular membership function of the ZR subset of error of a quantity equal to I_e, if it is larger than 10, or equal to Ie/4 if it is lower than 10.

The shapes of the other membership functions of the error are forced to keep the crossing points at a membership degree value of 0.5 moving the top angles of the membership functions of subset NS and PS, and consequently the top angles of the other membership functions.

At the same time the singleton membership functions of the ZP and ZN subset of the outputs are displaced toward the centre value 2048 of a quantity equal to $\frac{I_o}{4} + \frac{I_e}{2}$. The membership function of the other subset are rearranged in order to equally devise the spaces between 0 and the ZN singleton position and between the ZP singleton position and 4095.

If the distance between the ZN and ZP singletons positions is lower than 100, the correction factor is reduced to $\frac{I_o}{16} + \frac{I_e}{8}$.

The new controller set-up is tested by a new cylinder run, and hence a new controller set-up will be defined. The algorithm will run until satisfactory performance, as defined by the user in terms of accepted maximum error, will be reached.

The controller optimisation runs start from the completely retracted rod position. This working condition let the controller set-up to be well defined for the rod outwards motion, while the membership functions involved in inwards motion usually still need some corrections because of the chamber volume difference. Once the required performance are reached for outwards runs, some inwards runs must be carried out, applying the same algorithm only to the membership functions of the involved output quantities. Usually a maximum of three iterations allows the optimum shape detection.

The described algorithm was developed according to the formerly acquired experience in fuzzy controller tuning by trial and error method.

It drives the system to reach a suitable solution in a relatively low number of iteration (usually between 6 and 9) provided that the starting configuration is correctly defined.

The following figures show the evolution of performance and of the controller set-up during a tuning cycle.

Fig. 3 shows the evolution of the trajectory in the t-x plane, where x is the rod position. The reference position is 100 mm, i.e. middle course. The delay in motion start was imposed in order to allow an easy detection of the different plots. With the first attempt controller set-up the rod crosses the reference position, oscillates around a medium value and finally stops at a position higher than the reference one, with an error of about 10 mm, i.e. about 10% of the reference. The error and the oscillation width decrease at each iteration, until a satisfactory result is obtained at the seventh iteration. It can be pointed out that all the unsatisfactory set-up drive the rod through a trajectory characterised by a first high speed part, followed by an abrupt slow down, and later speed increases again when the rod crosses the reference position before beginning the oscillation around the medium value. This behaviour is due to an abrupt opening of the exhaust valve of the front chamber when the reference position is crossed. This undesirable effect is due to a bad positioning of the ZN and NS membership functions of the output quantities. The progressive displacement of the ZN and ZP singleton membership functions through the central position and the contemporary reduction of the width of the ZR subset of the error solve the problem and define a controller set-up which correctly drives the servoaxis.

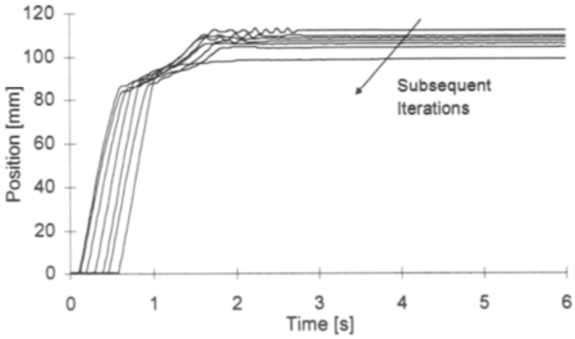


Fig 3 – Trajectory evolution trough iterations

Figure 4 shows the first and last iteration trajectory in the e-Δe plane in which the system performance is evaluated. Both e and Δe are represented with the value which are elaborated by the controller, according to (3) and (4).

The ellipses used to compute the performance indexes are drawn on the plots in order to point out how they can actually define the obtained performance with a good approximation and hence they can be used to compute correction to the controller set-up.

Because of the method the indexes are computed, it is not possible to detect a configuration in which the ellipsis reduces to a point. Maximum acceptable values of both the indexes must hence be defined before starting the optimisation algorithm.

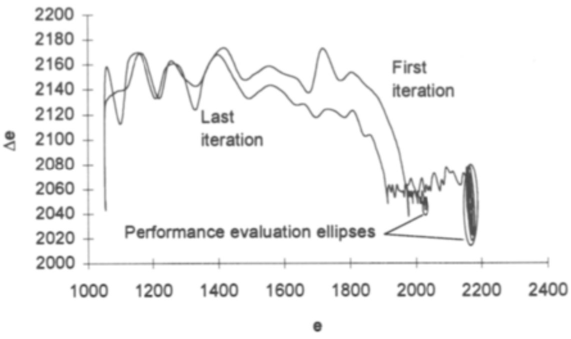
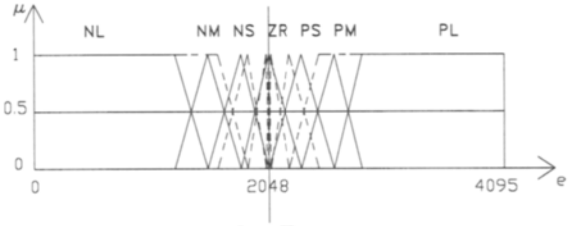
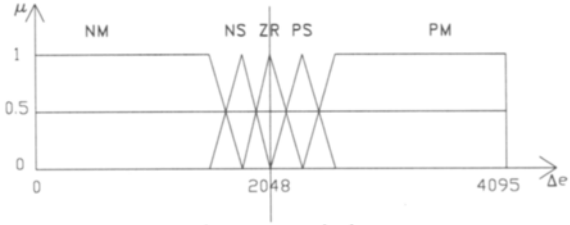


Fig. 4 – Trajectory in the e-Δe plane at the first and last iteration

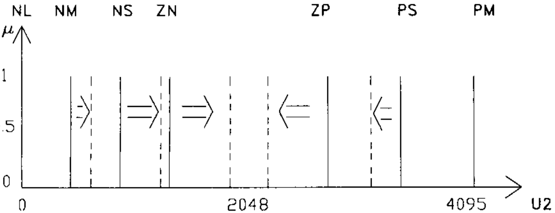
Finally fig 5 shows the evolution of the membership functions of input and output quantities. Particularly each figure shows the membership functions shapes at the first and last iteration.



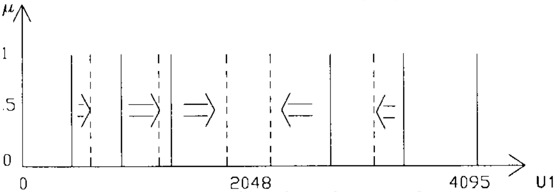
5a - Error



5b Error variation



5c – Front chamber duty cycle



5d – Rear chamber duty cycle

Fig. 5 – Evolution of membership functions from first to last iteration

5. CONCLUSIONS

A method for the research of optimal membership function for a fuzzy controlled pneumatic servoaxis was developed. The method applies methodically the experience acquired during trial-and-error tuning of several controllers for a family of similar servoaxis, characterised by the same layout and different dimensional parameters.

As the method is dedicated to a well known system the rules were formerly defined.

The algorithm operates on line, updating the membership function shapes at each iteration and directly testing their performance.

A method for performance evaluation was defined and used to compute two indexes whose elaboration allows membership functions corrections to be computed.

REFERENCES

1 M. Lopresti, R. Poluzzi, A.M. Zanaboni, “Synthesis of fuzzy controllers through neural networks”, Fuzzy Sets and Systems 71, pp. 47-70, 1995

2 A. Blanco, M. Delgrado, I. Requena, “A learning procedure to identify weighted rules by neural networks”, ”, Fuzzy Sets and Systems 69, pp. 29-36, 1995

3 Chuan-Tsai Sun, “Rule-Base Structure Identification in an Adaptive-Network-Based Fuzzy Inference System”, IEEE Transactions on fuzzy systems, vol. 2, n° 1, pp. 64-73, February 1994

4 Ching-Teng Lin, C.S. George Lee, “Reinforcement Structure/Parameter Learning for Neural-Network-Based Fuzzy Logic Control Systems”, IEEE Transactions on fuzzy systems, vol. 2, n° 1, pp. 46-63, February 1994

5 F. Pin, Y. Watanabe, “Automatic generation of fuzzy rules using the fuzzy behaviourist approach: the case of sensor based navigation”, Intelligent Automation and Soft Computing, Vol. 1, n° 2, pp. 161-178, 1995

6 Young Moon, Un-chul Moon, Kwong Y. Lee, “A Self-Organising Fuzzy Logic Controller for Dynamic Systems Using a Fuzzy Auto-regressive Moving Average (FARMA) Model”, IEEE Transactions on fuzzy systems, vol. 3, n° 1, pp. 75-81, February 1994

7 P. Viljamaa, H. N. Koivo, Tuning of a multivariable fuzzy logic controller”, Intelligent Automation and Soft Computing, Vol. 1, n° 1, pp. 15-28, 1995

8 G. Mattiazzo, S. Mauro, M. Velardocchia “Design of pneumatic positioner with PID and fuzzy control” - 6th International Workshop on Robotics in Alpe-Adria-Danube Region, RAAD’97, Cassino, Italy 26 - 28 june, 1997

9 Belforte G., Raparelli T., Velardocchia M. “Fuzzy Control of a Pneumatic Servosystem and Comparison with System using other Controls”, *Proc. of IFCA*, Terrassa Barcelona, 1993, pp. 712-718.

10 G. Belforte, G. Mattiazzo, S. Mauro, M. Velardocchia, T: Raparelli “Mechanical system with fuzzy control: design methodologies” - WILF ‘97 II Italian Workshop on Fuzzy Logic promoted by IEEE Neural Network Council, 19 - 20 march, 1997, Bari.

11 Mattiazzo, S. Mauro, M. Velardocchia, T. Raparelli “Articulated thumb for grasping devices” - The Fifth Scandinavian International Conference on Fluid Power, SICFP ‘97, Linköping, Sweden, 28 - 30 may, 1997, pp. 2-14.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

AN EXPERIMENTAL WORKBENCH FOR INNOVATIVE PNEUMATIC UNITS COOPERANT UNDER DISTRIBUTED CONTROL

E. Ravina[a] and R. Fattori [b]

[a]Department of Mechanics and Machines Design
Engineering Faculty – University of Genoa
Via Opera Pia 15A – I-16145 Genoa, Italy

[b]FESTO Italy
Via E.Fermi 36/38, I-20090 Assago (Milan), Italy

The paper describes the most relevant aspects of a research activity, still under development, concerning the performance analysis of advanced cooperant pneumatic units and components driven by different distributed control systems. Field bus, ASI and servo positioning strategies are applied to flexible and innovative devices.

1. INTRODUCTION

Competitiveness of modern automatized pneumatic systems is strongly related, from one side, to synergetic combination of advanced mechanical components with control electronic devices and, from another side, to the implementation of more and more diversified, flexible and modular plants. A wide spectrum of manufacturing industries requires more and more complicated units for handling, motion, assembling: a large number of digital and analogue pneumatic devices are involved, often operating under extreme dynamic conditions, high complexity automatic sequences, heavy working frequencies. The mechatronic machine design and the choice of the most suitable solution involves in particular problems related to the mechanical performances of pneumatic components and to their response under different control systems.

The development of control systems applied to industrial pneumatic automation makes available innovative solutions able to decentralize on field operation functions traditionally centralized. Conventional architectures based on point-to point connections aren't in many cases advantageous for high costs related to assembling, installation and maintenance, for high wiring errors probability, for limited expansion and flexibility. Microprocessor control units distributed along the plant and interconnected by communication network improve flexibility: according to different needs of the specific pneumatic application, "master–slave" structures can be organized in totally free way, involving active or passive devices like valves islands, I/O modules, video terminals, open interfaces, and so on.

Considering innovative digital pneumatic systems, the decentralization "on field" of specific operating functions assumes particular significance in presence of "intelligent" modules. The management of very different devices (mechanical actuators, sensors, advanced PLCs, industrial PCs, etc.) involves different entry and dialogue methodologies. Master-slave methods, with cyclic transfer of information, are combined to innovative communications, where supervisor modules organize the dialogue of smart terminals (local PLCs embedded on valves groups and on I/O terminals, in the same integrated island). Automatic machines can be designed strongly modifying and enlarging the outline of pneumatics, and following, for instance, the sub-groups logic, each of them able to autonomously manage specific operations. In addition, the most recent developments allow the control of scattered pneumatic units by means of advanced circuit logic (e.g. the new concept of Actuator-Sensor-Interface, designed as the most uncomplicated method connecting actuators and sensors to a controller), with electric power and communication signals transmitted together on a single bipolar wire.

2. NEW TRENDS IN PNEUMATICS

The most recent trend followed by manufacturers of pneumatic systems is strongly oriented to satisfy the user needs, not only in terms of low costs or good cost/performance ratios, but mainly creating "more utility" and better connection with specific needs required by single applications. Three focussed goals which must be in parallel performed are:

i) function integration;
ii) function differentiation;
iii) customization.

The achievement of the first goal requires high flexible products, with a wide spectrum of possible variations. With reference to pneumatic driving the function integration is realized by standard pneumatic actuators, by combination between actuators and mechanical guides, by integrated actuators for one or two axes (linear or linear and rotating). In the field of pneumatic distributors practical solutions can be found in valves manifolds, in units with field bus and ASI connections, in valves with "decentralized intelligence".

The research of more simplicity, with economical solution to driving and control problems through the use of product functions strictly related to the specific application, corresponds to the second goal. Considering, for instance, linear actuators, an interesting and practical answer to function differentiation consists on standard cylinders able to offer all the possibilities for fitting.

Finally, in order to satisfy special needs, the product functions are optimized for a specific use. This goal is performed optimizing the pneumatic device for particular tasks like cycle frequencies, payload, compactness, control of operating parameters and so on.

In any case the achievement of these results requires pneumatic components with high modularity and very flexible driving and control systems.

New problems arising in pneumatics concern the cooperant use of different pneumatic unit: more and more advanced pneumatic applications foresee servo-pneumatics integrated to digital modules driven by centralized or distributed control.

Modular robots, assembling cells, assembling cooperative systems, special machines (fitting units, electrical and electronic components assembling machines, ..), screwing, drilling or melding automatic stations, not standard automatic cells, packaging and food machines, units for sorting and storage of mechanical devices, control and test sub-systems are pneumotronic examples where different kind of pneumatic servo-systems must cooperate.

Starting from these considerations a specific research activity oriented to the analysis and identification of new pneumatic units cooperating under different distributed controls is implemented.

3. THE EXPERIMENTAL APPROACH

In order to give a contribution to the analysis of advanced pneumatic systems an experimental workbench, specifically designed for tests on co-operant pneumatic units and driven by different control systems, has been assembled at the Department of Mechanics and machine design of the University of Genoa in cooperation with FESTO Italy. Figure 1 synthesizes its highly flexible structure: an industrial personal computer is interfaced to cooperating stations, respectively controlled by PLC, ASI module and field bus. The main purpose of this experimental test bench is the analysis of problems of integrated control in pneumatics, simulating the presence of different and cooperating pneumatic subsystems and deepening the aspects related to their software synchronization and to the overall dynamic response, under actual working conditions.

Particularly interesting are original approaches, later described, related to monitoring and to on-line diagnostics of cooperating sub-stations. These topics, fundamental in a correct management of high complexity pneumatic systems, are faced making available flexible tools able to guide the user to detect the most probable causes of trouble or malfunction, in order to achieve systematic procedures of preventive and under condition maintenance.

Within each stations advanced and innovative pneumatic devices are under test; in particular:

i) compact valve terminals, combining high pneumatic performance with versatile communication. The decentralized concept allows on-site installation, in close proximity to the drive unit, optimizing the hoses lengths and response times. These

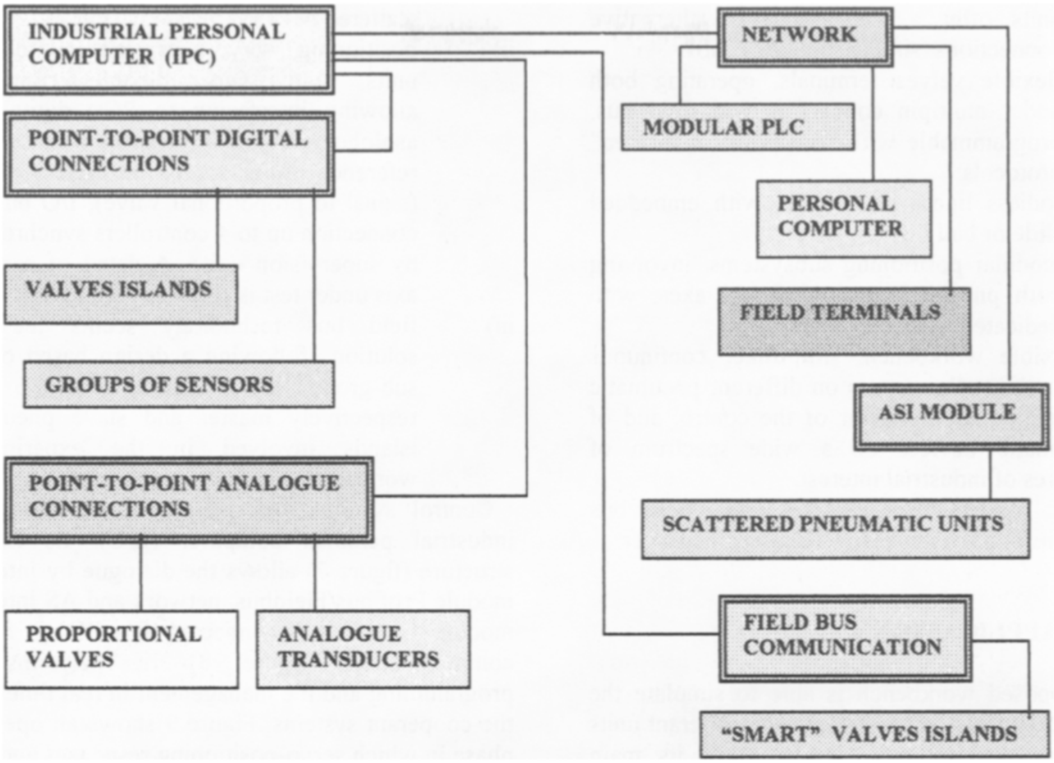

INDUSTRIAL PERSONAL COMPUTER (IPC)
NETWORK
MODULAR PLC
POINT-TO-POINT DIGITAL CONNECTIONS
PERSONAL COMPUTER
VALVES ISLANDS
FIELD TERMINALS
GROUPS OF SENSORS
ASI MODULE
POINT-TO-POINT ANALOGUE CONNECTIONS
SCATTERED PNEUMATIC UNITS
PROPORTIONAL VALVES
ANALOGUE TRANSDUCERS
FIELD BUS COMMUNICATION
"SMART" VALVES ISLANDS

Figure 1. The experimental workbench concept.

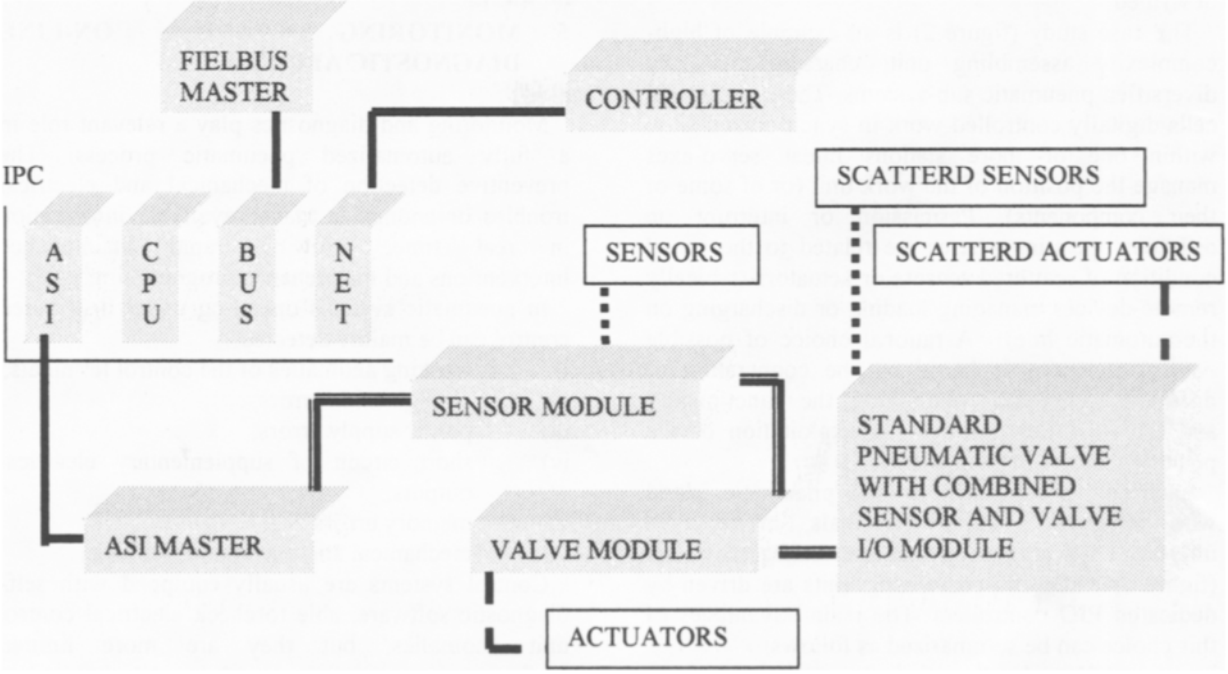

FIELBUS MASTER
CONTROLLER
IPC
SCATTERD SENSORS
A S I
C P U
B U S
N E T
SENSORS
SCATTERD ACTUATORS
SENSOR MODULE
STANDARD PNEUMATIC VALVE WITH COMBINED SENSOR AND VALVE I/O MODULE
ASI MASTER
VALVE MODULE
ACTUATORS

Figure 2. A case study.

units offer a wide range of alternative connections: simple, field bus, ASI;

ii) flexible valves terminals, operating both under multipin connection and field bus, programmable with "open" or "dedicated" protocols;

iii) rodless linear drive units with embedded slide or ball bearing guide;

iv) modular positioning subsystems, involving both pneumatic and electrical axes, with dedicated axis controllers.

The flexible workbench, still under continuous implementation, allows tests on different pneumatic units validating the behavior of the control and of the pneumatic devices on a wide spectrum of architectures of industrial interest.

Specific aspects regarding IPC and field bus applications in pneumatics are discussed hereafter.

4. AN APPLICATION EXAMPLE

The proposed workbench is able to simulate the actual behavior of a large quantity of cooperant units of industrial interest. In order to show its main characteristics and potentialities an application involving IPC field bus and servo-positioning axes is described .

The case study (figure 2) is an example of high-complexity assembling unit, characterized by diversified pneumatic sub-systems. Three pneumatic cells digitally controlled work in synchronized way: within one or more stations linear servo-axes manage the position of the work unit (or of some of their components). Permission or interrupt on different operating phases are related to the actual condition of scattered sensors or actuators (typically remote devices managing loading or discharging on the automatic line). A rational choice of possible control strategies is based on the cooperation of different techniques, optimizing the functionality and the flexibility through the exploitation of the positive features of each of them.

Each cell is controlled by one pneumatic island with integrated "smart" terminals, master-slave interfaced. Scattered devices are manager by ASI (figure 3) and servo-positioning units are driven by dedicated PID controllers. The main advantages of this choice can be summarized as follows:

i) ASI technology is more convenient than fieldbus (and than the traditional point-to-point PLC connection) in the control of scattered devices;

ii) positioning servo-axes are autonomous units, with very flexible controllers allowing interfacing to PCs, digital I/O, analogue inputs (position transducers, reference signals,...), analogue outputs (signal to proportional valve), I/O bus (for connection up to 4 controllers synchronized by supervision unit). A detail of a servo-axis under test is shown in figure 4;

iii) field bus technology seems the best solution following a design based on the sub-groups logic. Figures 5 and 6 show respectively master and slave pneumatic islands involved in the experimental workbench.

Control systems are interconnected through an industrial personal computer (IPC). Its flexible structure (figure 7) allows the dialogue by interface module Profibus/Fieldbus, network and AS interface module. This unit is connected to a PC via RS232 communication (figure 8): it performs the programming and the management in real time of all the cooperant systems. Figure 9 shows an operating phase in which servo-positioning responses and field bus programs are analyzed together.

5. MONITORING AND ON-LINE DIAGNOSTIC APPROACHES

Monitoring and diagnostics play a relevant role in a fully automatized pneumatic process. The preventive detection of mechanical and electrical troubles or anomalies makes available information in real time, allowing rapid maintenance interventions and maintenance programming.

In pneumatic systems operating under distributed control can be mainly detected:

i) working anomalies of the control terminals;
ii) configuration errors;
iii) power supply errors;
iv) short circuit of supplementary electrical outputs;
v) memory errors;
vi) mechanical and pneumatic troubles.

Control systems are usually equipped with self-diagnostic software, able to check electrical control unit anomalies; but they are more limited performances to recognize mechanical and electrical failures. Moreover the efficiency of a diagnostic system also depends on the number and the type of

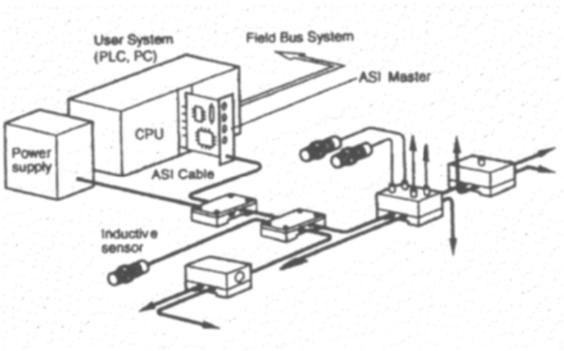

User System
(PLC, PC)
Field Bus System
ASI Master
CPU
Power supply
ASI Cable
Inductive sensor

Figure 3. The ASI concept.

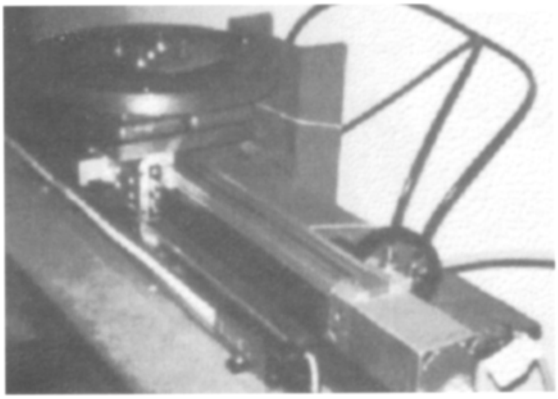


Figure 4. Servo-axis under test.

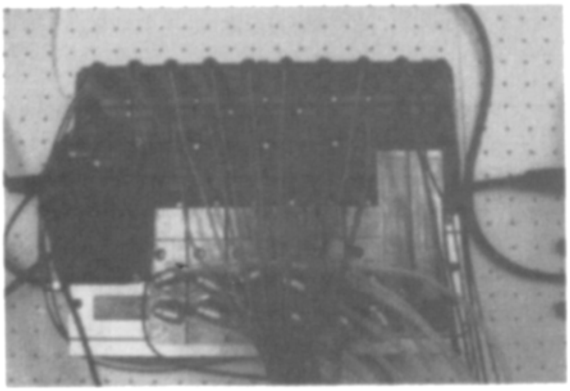


Figure 5. Field bus: valves island (master).

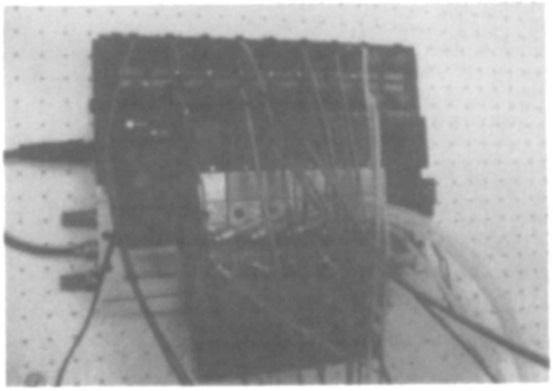


Figure 6. Field bus : valves island (slave).

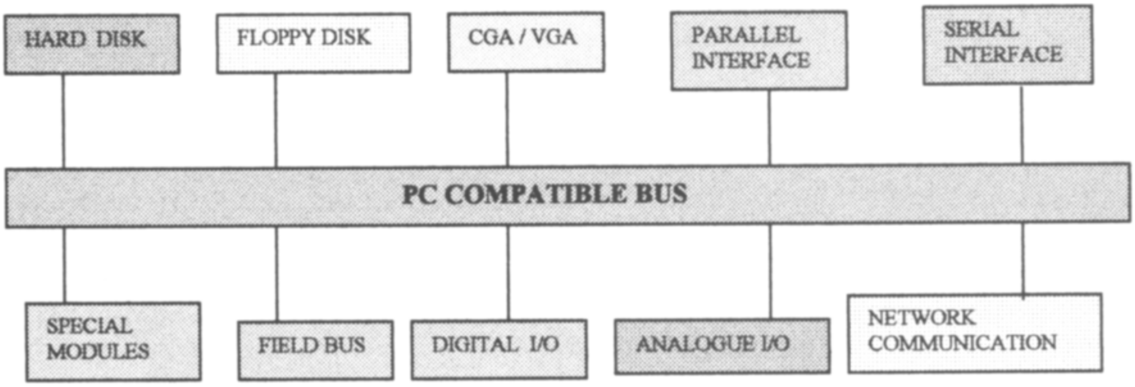

HARD DISK
FLOPPY DISK
CGA / VGA
PARALLEL INTERFACE
SERIAL INTERFACE
PC COMPATIBLE BUS
SPECIAL MODULES
FIELD BUS
DIGITAL I/O
ANALOGUE I/O
NETWORK COMMUNICATION

Figure 7. IPC architecture.

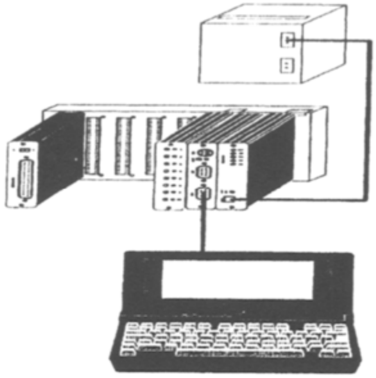
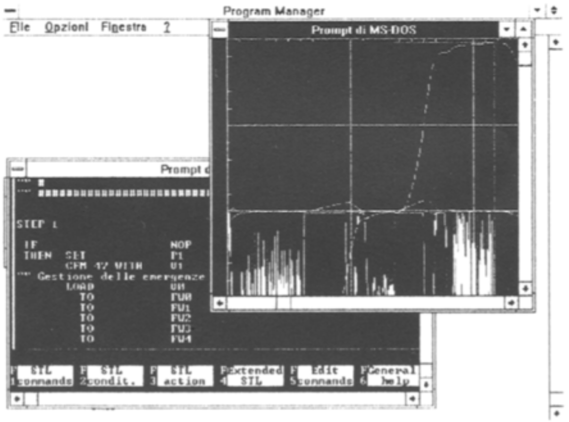


Figure 8. IPC/PC connection.



```
STEP 1

 IF                    NOP
 THEN  SET             M1
       CFM 42 WITH     U1
*** Gestione delle emergenze
       LOAD            UH
         TO            FW0
         TO            FW1
         TO            FW2
         TO            FW3
         TO            FW4
```

Figure 9. Example of concurrent analyses.

information made available to the user. For this reason the proposed workbench is customized with an on-line troubleshooting procedure, running on PC, and executable as concurrent procedure to other codes. Original checks located within the automatic sequence programs are able to compare actual responses to theoretical ones and consequently to activate warning messages or automatic interventions. The procedure is able to analyze almost 150 different anomalies or fault conditions, with automatic definition od possible corresponding causes and suggestions of possible remedies.

This approach has been in particular implemented on field bus units [10] and it is now under further development. Statistical analysis of errors can be added, in order to make available error trends and information about the most critical components.

6. SOME CONCLUDING REMARKS

The proposed experimental workbench seems to be an useful tool for acquire actual information on dynamic responses of advanced pneumatic components and sub-systems driven by cooperant control strategies.

Further deepening are oriented to optimize the integrated dialogue, enhancing the exchange of information in real time among each control system, under a supervision unit.

REFERENCES

1. T. Virvalo, Distributed motion control in hydraulics and pneumatics, Mechatronics, 1992 ,vol. 2, no. 3.

2. W. Bublitz, J. Pindel, P. Brunner, Integrating fluid power into control networks, Hydraulics and Pneumatics,1992, vol. 45, no. 5, p. 35-40.

3. P. Rohner, Modular programming concept of electronic controllers (PLC) for fluid power machinery. 2nd Intl. Conf. 'Fluid power transmission and control, Zhejiang, China, 1989.

4. J. Watton, J. Nelson, Mechatronic aspects of fluid power control. Mechatronics, 1993, vol. 3, no. 2.

5. S. Ferraris, A. Lucifredi, E. Ravina, Experiences of mechatronics applied to the design and management of oil-hydraulic and pneumatic systems. Intl. Conf. 'Mechatronics-The integration of engineering design', Dundee, Scotland, 1992.

6. S. Ferraris, A. Lucifredi, E. Ravina, Flexible oil-hydraulic and pneumatic circuit design for industrial automation, Intl. Conf. 'Industrial Automation', Montreal, Canada, 1992.

7. E. Ravina, Experimental mechatronics: an aspect of the motion control in fluid power systems" in Proc. of the Joint Hungarian-British Intl. Mechatronic Conf., Budapest, Hungary, 1994.

8. E. Ravina, Slot-type rodless pneumatic actuators: an experimental identification methodology based on acceleration analysis, in Proc. Intl. Congress IFPE 96, Chicago, USA, 1996.

9. E. Ravina, A diagnostic micro-lab for pneumotronic systems, 5th UK Mechatronic Forum Intl. Conf., Guimaraes, Portugal, 1996.

10. E. Ravina, On the experimental optimization and diagnosis of field bus pneumatic systems, IASTED Intl. Conf. 'Applied Modelling and Simulation', Banff, Canada, 1997.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Development of a new electronic-pneumatic control for pneumatic systems

G.Belforte, G. Eula, C.Ferraresi and V.Viktorov

Department of Mechanics - Politecnico di Torino Technological University
C.so Duca degli Abruzzi,24 - 10129 Torino - ITALY

This work presents the study, realisation and optimisation of an innovative electro-pneumatic interface that can be controlled with low power acoustic-electric signals. The analysis was made experimentally by checking which system geometry, and which type of command, improved the working of the device.

1.INTRODUCTION

The pneumatic systems have had more and more integration with sensors and with control electronic devices. In particular the study of interconnection between those systems and bus lines, capable of trasmitting electronic signals, requires the development of new interfaces controlled by low power electric control signals.

The disadvantage of the electrical commands normally used is the signal's voltage (generally 24V), the problem being that such a value in some applications is excessive. In addition to this, during their operation these systems absorb a certain amount of power. So there is a need to study the possibility of realising pilot stages with constructively competitive commands, capable of operating at supply voltages much lower than 24V, and which will also reduce the level of energy consumption.

Nowadays electrovalves with electromagnetic command realise the connection between electric and pneumatic signals. Usually to reduce the power required in the control signals, electromagnetic pilot stages are used. Therefore the minimum electric power are in general about 0.5-1 mW.

In many cases these values are easily available, but using direct commands, with electronic controllers or with bus lines, it is necessary to amplify signals before their use. Low power control electro-pneumatic valves have been already constructed, using piezoelectric pilot stages, but characterized by not too high dynamic behaviour: they are used especially in explosive enviroments because of their intrinsic safety characteristics.

Now there is an increasing interest in the development of low power and small dimensions interfaces, studying how to miniaturize pneumatic components without any reduction in their perfomances.

This work studies the possibility of developing a new kind of very low power electronic-pneumatic interfaces, controlled by signals (about few milliwatt) directly generated by bus lines at a voltage less than 5V. Among the numerous methods available, the transition of a fluidic jet from laminar to turbulent operating condition has been used. This transition were already used in sixty years to construct fluidic turbulent amplifier [1-3].

Nowadays it has been discovered again for the realisation of optopneumatic interfaces [4-6]. In a previous study of some authors, the feasibility of such an interface has already been verified [7].

Up to now the influence of the device's geometry and of an acoustic control signal about the transition of the jet from laminar to turbulent conditions has not widely analyzed in literature .

So in this work the turbulence amplifier's geometric characteristics have to be defined with the aim of obtaining at its exit the maximum difference of recovery pressure, from the moment when the control signal is active and the moment when it is not applied, and to choose the type of control capable of commuting the jet with the minimum use of power.

In the cases here considered the control signal may be generated pneumatically, acoustically and by the vibration of a piezoelectric element, while the inside geometry of the prototype can be modified so as to test different working conditions.

Figure 1 shows an example of circuit where this electrofluidic interface will be able to work. In this scheme the pressure reducer PR regulates the supply pressure of the electrofluidic interface E. Upon exiting, this element has to return pressure near to zero or at maximum possible recovery pressure, depending on whether or not the electrical control signal C is active. The ouput pressure from the fluidic element E is then sent to an amplifier A, which transforms this output signal into a pressure that is appropriate for the commutation of a standard pneumatic power valve V. The whole system indicated as PS is the pilot stage.

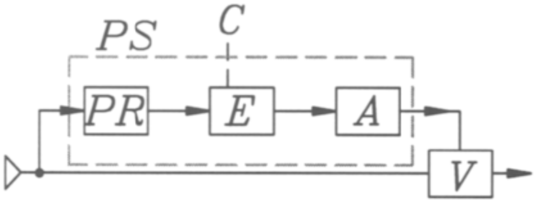
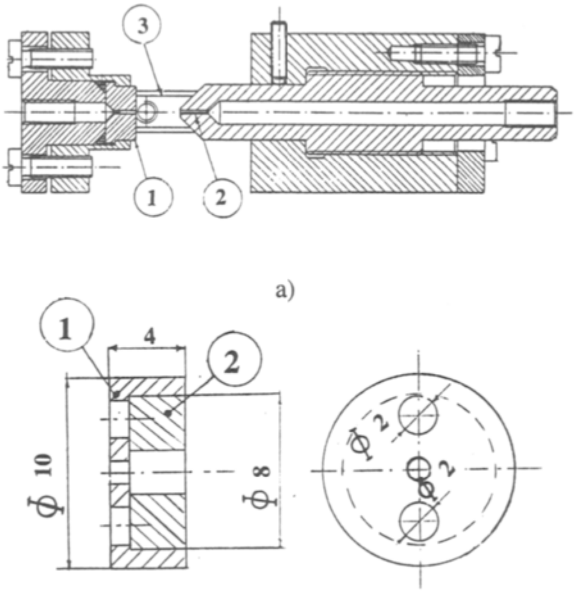


Figure 1. Application of the new interface.

2. DESCRIPTION OF THE PROTOTYPE

Laminar to turbulence transition is the basis of the operating principle of turbulence amplifiers: these are fluidic elements without moving parts, which use fluidodynamics whereby a main supply jet (laminar) is disturbed and made turbulent by the application of a control signal. As can be seen from the literature, the transition of fluid from a laminar regime to a turbulent one is influenced by many factors such as, for example, speed and viscosity of the fluid, shape and sizes of the cross section of the duct in which the fluid flows. To ascertain the way in which the geometry plays on the laminar to turbulent transition of a jet of air, a prototype was made in which the most important geometric parameters could be varied. This prototype is shown in figure 2a. It is made of a body inside of which two nozzles have been positioned, 1 and 2, flanked on a chamber 3 that is open on two sides in order to permit the discharge of air. Indeed the latter, coming from the supply (nozzle 1) is, in the absence of the control signal, laminar and is gathered at the exit (nozzle 2) with a very high retrieval level. The presence of the control signal provokes turbulent transition and a fall of the exit signal, because in this case the jet becomes much larger and becomes dispersed on the outside via the openings of the chamber 3. The operation of the device is improved the more the signal tends towards zero at the moment in which the control signal is applied. Furthermore, in order to have a laminar flow, there is a need for a very low supply pressure and a very restricted cross section of the emitter nozzle (typical values are a nozzle diameter of 0.25 mm and pressure equal to a few thousandths of Pascal). The emitter nozzle 1 is in this case made in two parts, as it would be difficult to make a hole of 0.5 mm in diameter concentric with the connecting line should the nozzle and the fastening flange be one single body.



a)

b)

Figure 2. a) Laminar-turbulent amplifier prototype. b) Diffuser element.

Furthermore has been made with just one fastening flange to test out nozzles with differing geometries. The absence of leaks between nozzle and flange is guaranteed by gasket a made in such a way as to ensure the seal without altering the geometry of the supply duct.

As suggested in previous studies on turbulence amplifiers [8][9], an auxiliar device has been placed between emitter and receiver to improve the system's performance. This element, known as the diffuser, is made up of two coaxial discs (1 and 2 in figure 2b) with axial symmetry and circular in shape, having central holes with different diameters.

Furthermore, the measuring device has the following features: the dimension of the chamber 3 that exists between the nozzles does not disturb the jet and presents sufficiently large discharge ways for a rapid outflow of air not recovered by the receiving nozzle 2; it can send, via a purpose made conduit, a pneumatic control signal supplied separately by the emitting nozzle 1; it is possible to send an acoustic control signal, as the chamber 3 between the two nozzles is open and sound waves can hit the jet and generate the laminar to turbulent transition; the shape of the above-mentioned chamber 3 allows for the insertion, at variable distances, of elements between the nozzles 1 and 2; the device guarantees the coaxiality of the two nozzles 1 and 2 regardless of the distance at which they have been placed.

3. TEST METHODOLOGY

The experimental tests were made using different circuits, depending on the sizes that had to be obtained. The analysis was made by varying the geometry, the supply pressure, the supply flow, the type of control signal and its intensity. From a geometric point of view the following were varied: the angle α of the emitting nozzle (30°-45°-60°), the angle β of the receiving nozzle (45°-60°) and various distances L (6-16 mm) between the nozzles, with the aim of identifying the optimum geometry for the jet's laminarity. The above-mentioned parameters are illustrated in the scheme of figure 3. The supply flow rate was measured to calculate the Reynolds number, defined as:

$$\mathrm{Re} = \frac{\rho v d_s}{\mu} \qquad (1)$$

where v= velocity of the fluid; d_s= characteristic geometrical dimension (supply nozzle diameter); ρ= density of the fluid; μ= viscosity of the fluid.

The first control signal used was a pneumatic one with variable values for the control pressure from 0 to 1000-1500 Pa; the second was acoustic (with variable frequencies from 2000 to 22000 Hz and intensities of 90-100-110 dB), the third was electrical, using piezoelectric devices supplied with varying voltage from 0 to 20 V. In general the supply pressures were set between 2000 and 3000 Pa. Tests were also made in the complete absence of any control whatsoever, in order to evaluate the value of the pressure recovered on the receiving nozzle under such conditions. Finally, the diffuser shown in figure 2b was used to see if it was possible to reduce the level of constant pressure which is permanent at the exit even after the application of the control.

4. ANALYSIS AND RESULTS

4.1. Tests using different geometries and a pneumatic control

Figure 3 shows the test bench used to analyse the behaviour of the prototype when the main geometric parameters were varied and when a pneumatic control was used. In this scheme the prototype is represented by two nozzles 1 and 2, while nozzle 3 creates a pneumatic control. The working fluid (air) enters the system via supply 4, and resistance 5 enables the regulation of supply pressure. The manometers 6, 7, 8 measure the supply pressure (p_s), output pressure (p_o) and control pressure (p_c) respectively. The latter is regulated using resistance 9.

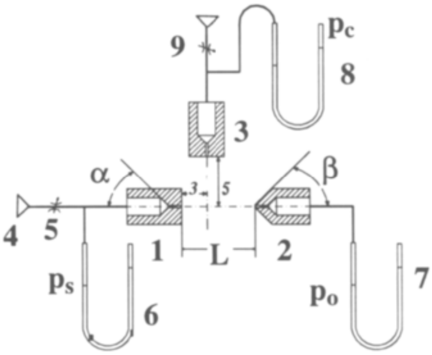


Figure 3. Experimental set-up used varying geometry and with pneumatic control.

The influence of the inclination of the connecting area between the supply duct and the emitting nozzle was then analysed: the latter has a circular cross section with a 0.5 mm hole, while the supply channel has a diameter equal to 4 mm. It is therefore necessary to have a conical section which joins together the two holes, the inclination of which can influence the operation of the device. Three nozzles with different values for taper angle α (figure 3) were therefore built.

The shape of the receiving nozzle was also studied to check its effect on the laminar to turbulent transition of the jet, and on the value of the retrievable output pressure. With this in mind, different geometries with different values for angle β (figure 3) were constructed The influence of distance L (figure 3) between the nozzles was studied by either placing these further apart or bringing them closer together, in a field in which it was possible to observe the phenomenon of the laminar to turbulent transition. The possibility of varying this distance is ensured by the special shape of the receiving nozzle 2, which can be moved along its own axis thanks to the threading made on its external surface. The minimum distance L used is equal to 6 mm, the maximum is 16 mm.

The first tests, carried out by varying both α, β and L, made it possible to draw a diagram of the output pressure p_0 as a function of the Reynolds number. In this way it was possible to choose the values for α and β that were most favourable to the laminarity of the supply jet. Following this, tests were made with a pneumatic control signal using as angles α and β the optimum values previously identified (α=45°, β=45°), while distance L was still varied (6-8-10-12 mm).

This enabled the identification of the geometry most suitable for obtaining a laminar jet which, as such, requires less power of the control signal for its commutation.

Figure 4 shows what happens to output pressure when p_c is varied, for different values of the supply pressure and for a geometry in which α=45°, β=45°, L=12 mm. The results were among the best obtained as one can see that low values for p_c are enough to obtain a commutation of the jet and therefore a rapid falling of p_o.

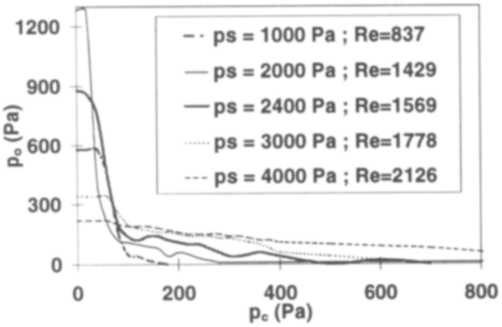


Figure 4. Characteristic p_o versus p_c with various p_s.

Analysing figure 4 one notices, however, the presence of a certain residue pressure at the exit of the element, even for high values (100 Pa) of p_c. Optimal supply pressures are about 2000 Pa because with p_s=3000-4000 Pa the device already works in turbulent conditions and its behaviour is not the optimal one. Instead in the case of p_s=1000 Pa the supply pressure is not enough for the right functionament of the system. Residue pressure is about 150 Pa for p_s=2000 Pa with p_c=100 Pa. In order to eliminate this inconvenience an element (diffuser) was placed between the two nozzles in order to deviate sufficiently from the axis of the system the lateral parts of the jet and thus reducing (or annulling) p_o in the presence of a command signal p_c. To this aim both the geometry of the diffuser and its distance from the emitting nozzle were seen to be important.

Initially between the two nozzles a disc having a central hole coaxial with the conducts of the nozzles of a diameter of 0.6, 0.7, 0.8 mm respectively has been experimented. The axial length of the element under examination is equal to 4 mm. Figure 5 shows the set-up used.

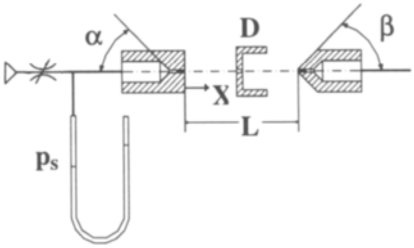


Figure 5. Optimisation of the system with diffuser.

The tests were carried out with α=30° and 45°, β=45°, L=10 and 12 mm, while distance X, between the disc and the emitting nozzle, was varied from 0 to 9 mm. The best results were obtained with a disc with a 0.7 mm hole and a distance of X=5 mm from the emitting nozzle. To diminish further the residue pressure, small discs were made to be inserted in the above-mentioned one, having central holes equal to 0.8, 1.0, 1.5, 1.8, 2.0, 2.2 mm. The new element obtained is called diffuser. Figure 6 shows what happens to p_o when p_c is increased, in the two cases in which between the nozzles of the system, either there is no object or the diffuser is placed. These results have been obtained with α=45°, β=45°, L= 12 mm, diameter disc 0.7 mm, cylindric diffuser with inside diameter hole 1.5 mm, X=5 mm, p_s=2400 Pa. In these conditions one obtains a high recovered pressure in the absence of a control signal, and a residue pressure of 50 Pa for control pressures above 100 Pa.

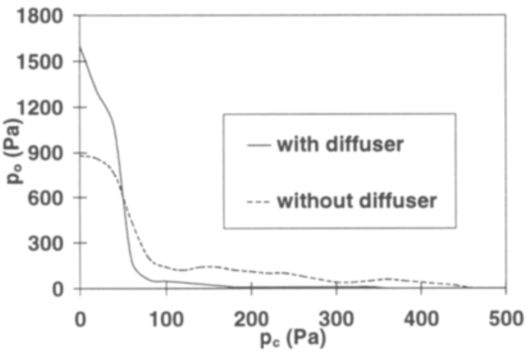


Figure 6. Influence of diffuser on p_o versus p_c.

It is therefore evident that the presence of the diffuser brings down the residue pressure and favours a better operation of the system. It was also seen that once the commutation had taken place, the difference between the residue pressure obtained for different values of p_s was minimal: so the system does not feel small variations of the supply pressure.

4.2. Tests with the acoustic control

Using the prototype with a geometry in which α=45°, β=45°, L=10 mm, and a diffuser with 0.7 and 1.5 mm diameter respectively, the test scheme illustrated in figure 7 was set up. In this case the command signal is obtained from sound waves generated with the loudspeaker L, connected to the amplifier A and to the signal generator G. The spectrum analyser AS reads the signal of the generator G and of the microphone M, amplified by the unit P.

These tests established how p_o performed when the acoustic signal generated by the loudspeaker was varied. A preliminary analysis, in particular, of the frequency response of the latter indicated that for frequencies between 2000 and 22000 Hz the amplitude emitted by the loudspeaker (measured in dB) remains constant.

Following this, one proceeded to the setting of the loudspeaker's supply voltage in such a way that the amplitude of the acoustic signal (dB) took on set and constant values. By varying the frequency of the loudspeaker signal, it was established for which values this produces the laminar to turbulent transition of the jet, or a notable reduction of p_o.

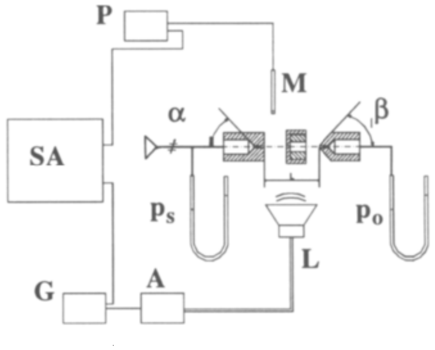


Figure 7. Set-up used with non-pneumatic control.

Figure 8 shows what happens to p_o as a function of the frequency f for different values of the acoustic intensity of the signal (dB), for a p_s=2000 Pa. From the graph it is clear that with an acoustic intensity of 110 dB, there is a strong reduction of p_o (about 1000 Pa) for frequencies between 7000 and 8000 Hz. The use of an acoustic command signal must therefore foresee to work in this frequency range if one wants to obtain a rapid and immediate reduction of p_o.

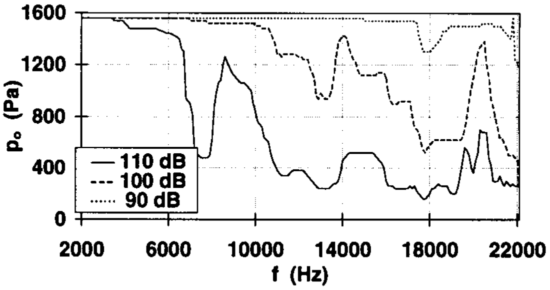


Figure 8. Characteristic p_o versus frequency.

4.3. Tests with control generated by piezoelectric elements

The test scheme used is similar to that of figure 7, where now, however, L stands for a piezoelectric element that generates an acoustic signal as a result of the deformation of a piezoelectric lamina. The tests enabled to evaluate what happens to the p_o when the supply voltage of the piezoelectric element was varied, using the following geometry: α=45°, β=45°, L=10 and 12 mm, d= diameter of piezoelectric element = 20 and 27 mm. The last parameter is important because previous studies [9] demonstrated the influence of this geometry on the performance of a turbulence amplifier. In fact the presence of the piezoelectric element changes the conditions on the amplifier's output ports, while its distance from the chamber between the two nozzles influences the control signal generation. Depending on the distance L between the nozzles, and the diameter d of the piezoelectric element, different fields of frequency were identified in which it is possible to obtain a reduction of p_o. This is illustrated in Table 1.

Table 1
Fields of frequency in which there is reduction of p_o

L (mm)	d (mm)	f (kHz)
12	20	4.96-5.22 and 9.51-9.56
10	20	5.1-5.3 and 9.5-9.6
12	27	16.8-17.1 and 21.9-22.1
10	27	17-17.2 and 21.7-22

The best results were obtained with α=45°, β=45°, L=12 mm, f= 5 kHz, p_s= 2400 Pa, d= 20 mm (figure 9). The tests were carried out keeping the frequency constant and varying the sonic intensity by varying the supply voltage of the piezoelectric element.

Figure 9 illustrates what happens to the output pressure when the voltage is varied. It is clear that a variation of very few volts (0 - 0.6 V) produces a notable variation of p_o (1400 Pa). The various curves refer to the three different supply pressures.

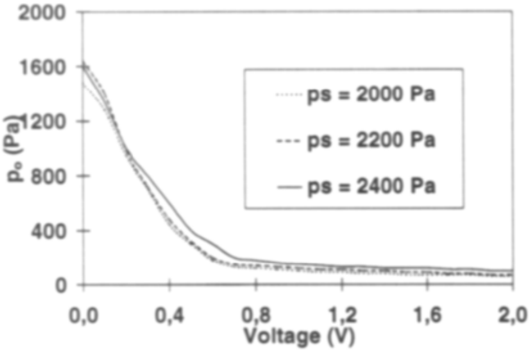


Figure 9. Output pressure obtained with different voltages of piezoelectric control and different supply pressures.

This is owed to the fact that the closer the working frequency of the piezoelectric element gets to that of the jet, the more the jet becomes disturbed and the output pressure drops.

5. CONCLUSIONS

The best technical characteristics identified of the system presented here can be summarised as follows: geometry: α=45°, β=45°, L= 10-12 mm; diffuser diameter 0.7 and 1.5 mm respectively, X=5 mm; supply pressure: p_s =2400 Pa; maximum variations of p_o: using pneumatic control $\Delta p_o \cong 1540\ Pa$ with $\Delta p_c \cong 80\ Pa$; using acoustic control $\Delta P_o \cong 1000 Pa$ with frequencies between 7000 and 8000 Hz and amplitude equal to 110 dB; using piezoelectric control $\Delta p_o \cong 1400\ Pa$ with $\Delta V \cong 0.6V$.

Then the prototype under examination enabled the construction of an innovative pilot stage, controllable with electric signals. The results obtained were good in that by choosing the suitable geometry of the element, the supply jet can be kept laminar, thus making a device the commutation of which requires a minimal amount of power. The exit of the turbulence amplifier made can be amplified and then sent to common stages of pneumatic power. In this way the prototype built becomes an effective electro-acoustic-pneumatic interface.

ACKNOWLEDGEMENTS

We would like to thank Eng. Massimo Buratto for his collaboration during the tests.

The research was financed by MURST.

REFERENCES

1. Auger R.N., Fluid. Ampl. Symp., HDL, Whashington, USA 4 (1964).
2. Verhelst H.A.M., 2nd Cranfield Fluid. Conf., B.H.R.A., Cranfield, UK paper F2 (1967).
3. Siwoff F., 3rd Cranfield Fluid. Conf., B.H.R.A., Cranfield, UK paper H2 (1968).
4. Rolland J., 3rd Cranfield Fluid. Conf., B.H.R.A., Cranfield, UK paper T1 (1968).
5. Ferhmann R., Matko J., Shubert W., Shulze W. and Tittle H., 4th Cranfield Fluid. Conf., B.H.R.A., Cranfield, UK paper A2 (1970).
6. Hu F.Q., Watson M., Page M. and Payne P.A., Mechatronics Perg.Press, 3 No.3 (1993) 369.
7. Hu F.Q., Watson M., Page M. and Payne P.A., Applied Optics Digest-Pages, (1990) 97.
8. Bramley H.C., Hu F.Q.and J.M.Watson, European J. of Fluid Power, Noregren Martonair Limited, (1993) 16.
9. Belforte G., Ferraresi C., Eula G., 5th Triennal Symp. on Fluid Control, Measur. and Visual., Flucome'97, Hayama, Japan (1997) 551.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Smart Components-Based Servo Pneumatic Actuation Systems

J Pu, P R Moore, C B Wong and J H Wang

Mechatronics Research Group
Faculty of Computing Science and Engineering
De Montfort University
Leicester, LE1 9BH UK

Pneumatic drives are the most widely used actuation systems. However, research in relevant areas is very limited indeed, often carried out, if any, in isolated environments. There is a lack of strategic thinking and tools which can enable the development to go forward with a fruitful end. In recognising the advancements in PC-technology, field-bus systems, microprocessors and computing technologies, the authors advocate the concept of smart components-based servo pneumatic actuation systems. The paper will expand on the components-based nature of pneumatic drive systems to identify the templates for smart valves, smart sensors, smart actuators, etc.. The development of components-based pneumatic drive systems will be considered in the context of a recently closed UK EPSRC-funded research project on high speed pneumatic actuation, addressing issues such as simulation, algorithm development, sizing, tuning, installation & commissioning. A high speed pneumatically-driven gantry pick-and-place positioning system with novel design features will be studied to demonstrate the capability and advantageous attributes of pneumatic drives for sophisticated applications.

1. INTRODUCTION

Servo pneumatic actuation is now a recognised technology as an alternative to their electric and hydraulic parts in areas of high speed medium power application [1, 2]. However, an industrial uptake is not yet happening. There are a couple of important issues which needs to be addressed. Firstly pneumatic servo products needs to designed and packaged in a form which are user friendly, ideally perhaps with a use pattern which are familiar with engineers dealing with servo drives. For example, if a pneumatic servo can be tuned as a PID controller for electric drives, the users would feel much more confident and comfortable. Recent development in PC technology and fieldbus systems should influence significantly the way in which pneumatic drives can be used and packaged.

The potential market for pneumatic servos can be significantly huge if the units can be packaged in readily-use form which do not require much tuning, commissioning and maintenance.

2. COMPONENTS-BASED PARADIGM

To illustrate the components-based nature of pneumatic drive systems, Figure 1 depict a high speed pneumatic pushing mechanism which requires some level of profiling capability.

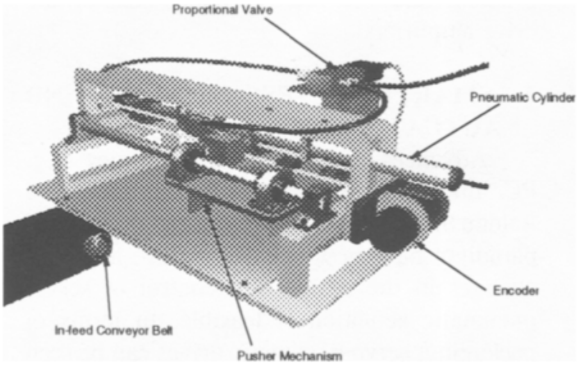


Figure 1 The Construction of A Pneumatic Pushing Mechanism

Figure 2 represents a components-based modelling paradigm adopted, in which pneumatic servo systems are decomposed into individual basic components [3].

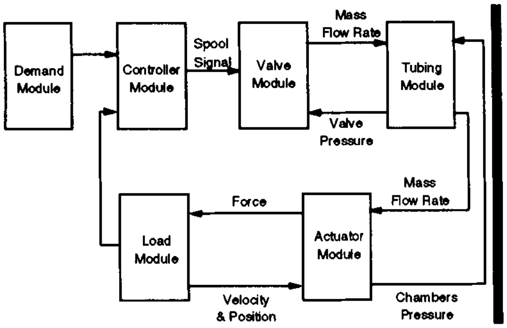


Figure 2 Components-based Modelling of Servo-pneumatic Systems

Comparatively speaking, pneumatic drives tend to be more modular construction and components oriented. For example, the amplifier for driving a pneumatic proportional spool drive is typically in the order of one ampere or less. Pneumatic actuation has a self-cooling effect as well. Potentially, a pneumatic servo system can be built as a single unit, where the actuator (cylinders or motors), the valve and the amplifier and the controller can be integrated in a modular physical form [4]. However, it is difficult to see how this can be achieved with electric counterparts because of the heat dissipation problem and the relative bulky size of the drive amplifier.

3. SMART VALVES, SENSORS AND ACTUATORS

Advances in enabling technologies (e.g. PC, microprocessors, digital electronics and fieldbus), together with the components-based paradigm have now reached a stage at which changes in the design and control of servo-pneumatic actuation is feasible. In terms of packaging, servo pneumatic drives can be seen to consist of four functional units, namely: sensors, actuators, valves and controllers. Intelligence can now be directly embedded into the sensors, valves, actuators and I/O components [5]. This type of new generation of 'smart' components can provide some informed rather than raw data, e.g. self-monitoring and diagnostic functions [6].

- Information stored concerning the device itself
- Some data concerning the life history of the device (life cycle data acquisition)
- Component validation
- Programmable operating characteristics

- *Smart Valves,* which may provide information concerning the valve itself, e.g. flow capacity, bandwidth, deadband, allowed operating pressure. A smart valve should have a build amplifier circuit. The flow characteristics of the valve may be made programmable. The valve may provide informed operating history about its life span.
- *Smart Actuators,* which can have either a built in feedback sensor system with digital signal which can be directly used by the controller. A smart actuator may also have built in integral control valves. Again a smart actuator may provide information concerning itself and its life history.
- *Smart I/O devices,* which may provide extended functions as those with Honeywell SDS devices.
- *Smart sensors,* which may contain self-validation algorithms and digital output to the controller
- *Smart pneumatic servo units,* in which the valves, the sensors, I/Os and the actuator are integrated as a single unit. This can be effectively achieved as previously discussed with pneumatic drives.

4. DEVELOPMENT TOOLS FOR PNEUMATIC POSITIONING SYSTEM

Software tools which support the design, simulation and control of servo-pneumatic drive systems are almost totally absent in the commercial market, perhaps largely due to the situation that servo-pneumatic drive systems are still not widely utilised in industrial applications. Design and control paradigms with a supporting software toolbox are advocated for developing high-speed servo-pneumatic actuation systems.

4.1. Modelling & Simulation

Matlab/Simulink environment is chosen to implement the adopted components-based paradigm. The individual component is implemented as a 'block' within the Matlab/Simulink environment. Defining a servo-pneumatic system is much like drawing a block diagram. Blocks are copied from the block libraries which is supplied either by model designer or component manufacturers.

During the derivation of the mathematical model for individual components, different levels of complexity can be achieved (The level of complexity is dependent upon the accuracy requirements, modelling efficiency and time constraints). Each component category consists of a collection of components bearing the same linkage specification in the component interfaces. A complete system is formed by selecting individual components from different categories and connecting them through their external interfaces (see Figure 2). Thus a components-based framework allows the addition and extension of any model irrespective of its sophistication requirements. The advantage is that some components can be modelled as closely as possible to the physical component (e.g. the most critical components), while others can be simplified to increase the modelling efficiency. Furthermore, empirical component models can be mixed with theoretical component models to form an overall system.

4.2. Design Tool Box Architecture

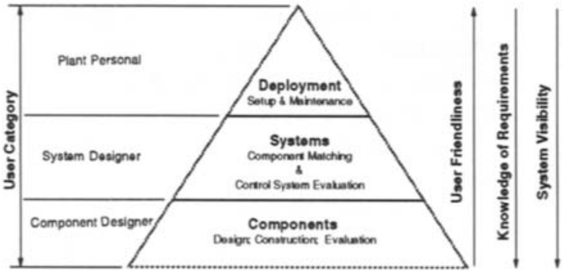


Figure 3 Conceptual Framework for the Pneumatic Design Tool Box

The software framework is in a hierarchical structure with 3 different layers to target different category of users.

- Components layer

The components layer provides the basic building blocks (components) to design and evaluate a servo-pneumatic system. Component models are created, evaluated and validated separately. The individual component is implemented either as a Simulink block or S-function.

The system segregation allows a reduction in model complexity and an increase their robustness which leads to faster model formulation. Ideally, component models are supplied by the component manufacturers because they have the first hand knowledge about their own components. Other sources to obtain the component models could be either from component designers (modellers) or experimentally (e.g. empirical models).

- Systems layer

The systems layer builds upon the component layer. A complete servo-pneumatic system model is built using the component oriented approach (see Figure 1). Different control strategies can then be evaluated without the real system having to be built.

- Deployment layer

The deployment layer is a configurable user interface that encapsulates all the system complexities. Only the tool (e.g. system tuning) that is to be used to set-up and maintain the system is exposed to the user.

4.2.1.User Interface

A configurable user interface developed with Microsoft Visual Basic is used to encapsulate the system complexities. A text file is used as a repository to house the configuration information. The configuration information is used to define the functionality exposure. The configurability of the user interface allows the system builders to customise a system commissioning and maintenance tool which is specific to the system/application type, without needing to re-writing the user interface code.

4.2.2.Components Sizing and Matching

The performance envelope of a servo-pneumatic system is largely depend upon on two key components and they are closely related namely, the control valve and the actuator. A tool that can be used to assist the designer/builder to size and match these two components is strategically important.

The components sizing and matching module is part of the design tool box and designed to integrate directly with the modelling/simulation environment. After completion of the selection process, the module will invoke the simulation environment to perform the performance evaluation. The evaluation result can be used to indicate how well the selected components match with the specifications.

The sizing and selection process is not a fully automatic process and requires human interaction. The main reason is the initial sizing and matching between components are mathematically based. Although the sizing and matching routine selects the best solution, these components might not be commercially available. Human interaction is necessary to check the chosen solution and to confirm the commercial availability.

5. Controllers and Algorithms Development

Current commercial motion controllers do not employ control schemes (control terms/algorithms) which are specific for controlling pneumatic drives. Motion controllers which can be used to control pneumatic drives are only available from very limited sources (e.g. Festo and Bosch). These controllers are typically application specific and in some cases can only be used with one actuator package. System and controller configurability are typically very limited. Essential features of the new control scheme utilising a basic PID plus velocity feed forward kernel can be summarised as gain scheduling; acceleration feedback; offset compensation; valve characteristics programming; time delay compensation; and position accuracy compensation. A full control scheme for servo-pneumatic systems was designed and implemented using the NEXTMOVE motion control platform from Baldor-Optimised Control. The NEXTMOVE is a 32-bit DSP based multi-axis motion controller which can co-ordinate up to 8 axes with auxiliary I/O support [7]. This controller with the modified control scheme can now be configured for use with either electric, pneumatic, or hydraulic actuators or a combination thereof.

A sub-set of the full control scheme for servo-pneumatic systems was implemented by Baldor on the SERVO-NODE single axis motion controller for point-to-point motion control. The SERVO-NODE is a 16-bit low cost platform, with appropriate interfaces to control pneumatic actuators with either analogue or encoder feedback.

A modular single axis motion control with fieldbus interfaces was developed by Baldor during the later stages of the research. This very compact, low cost platform known as the MME (Mint Motion Engine) enables the controller interface to be configured specifically for any actuator. Furthermore, because of the miniaturised hardware it can be embedded within the actuator body. In this way fully modular distributed actuators can be produced which can be linked via the fieldbus.

6. CASE STUDY: *A HIGH SPEED PNEUMATICALLY-DRIVEN GANTRY PICK-AND-PLACE POSITIONING SYSTEM*

A case study which based on the packaging of confectionery products in Mars Confectionery (UK) Plc. has been carried out. A generic two axes gantry type packaging machine has been built to demonstrated the high speed characteristics of servo-pneumatic actuation systems. The key features of the machine are.

- Low moving mass because of the direct-drive pneumatic actuation.
- High acceleration/deceleration - over 3g
- High Speed - 3.0 m/s (see figure 5)
- Payload - up to 2 kilograms
- cycle time - up to 40 cycles per minute

The machine consists of two axes of motion which are powered by pneumatic cylinders. One axis provides the swing action which allows the pick-and-place axis to move from the in-feed conveyor to the packaging container. The main axis is responsible for picking the product from the in-feed conveyor and placing it in a container for packaging (see figures 4 and 6).

The size and weight of the product varies dependent upon the production requirements. The stroke movement required by the pick-and-place axis varies during the stacking cycle. This application is a perfect example to demonstrate the high-speed characteristics of servo-pneumatic actuation systems and the control strategies adopted in our motion control scheme.

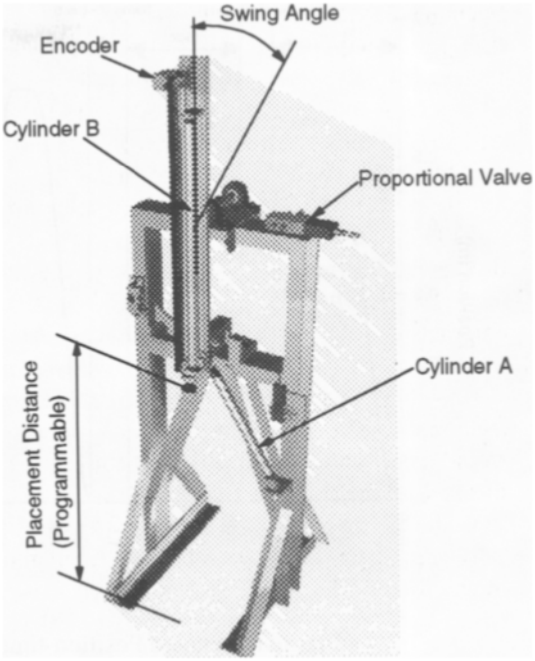

Swing Angle
Encoder
Cylinder B
Proportional Valve
Cylinder A
Placement Distance (Programmable)

Figure 4 Prototype Rig for the Pick and Place Packaging Machine (High-Speed Point-To-Point Motion)

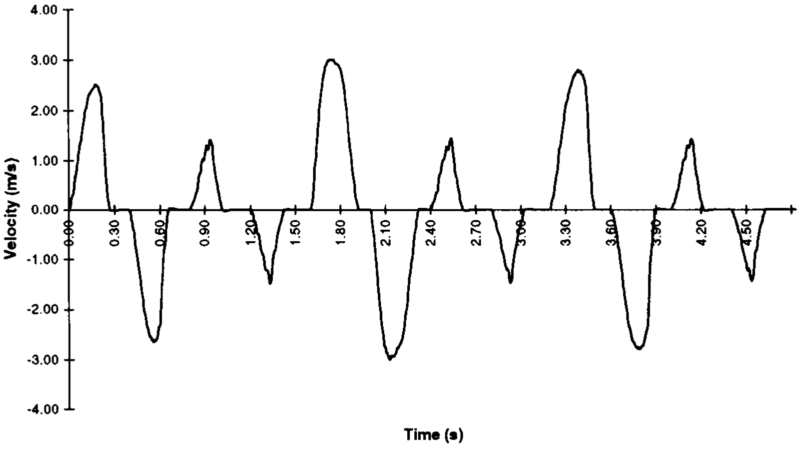


Figure 5 Actual Velocity-time graph for the pick-and-place cylinder

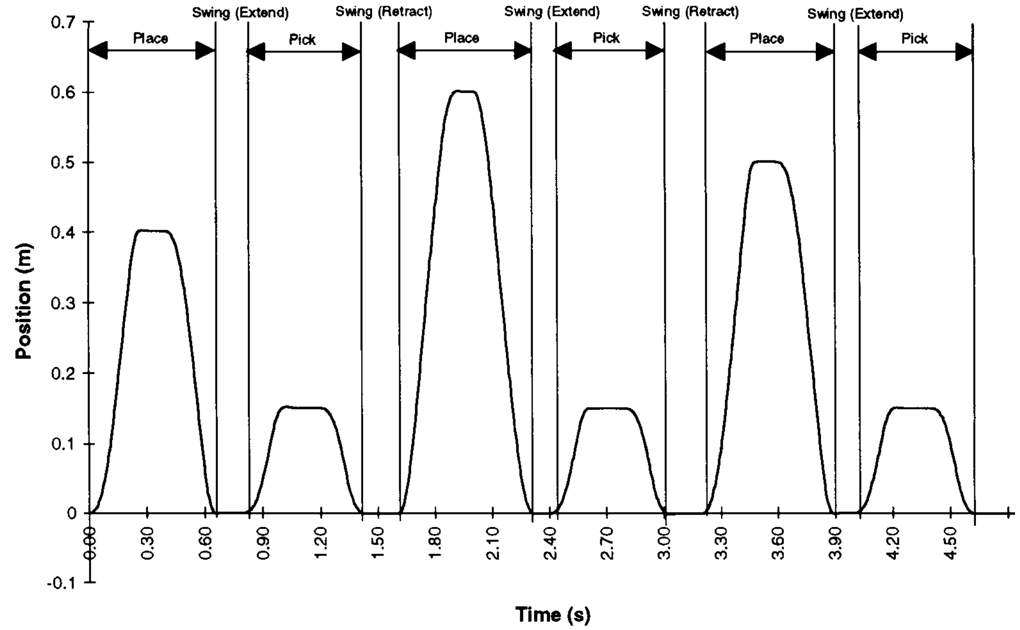


Figure 6 Demand Position-time graph for the pick-and place cylinder

7. CONCLUSIONS

Smart components-based servo pneumatic actuation system is yet to be fully developed to penetrate the market place. The research to date has shown that it is entirely feasible to produce a range of pneumatic servo actuation production with the inherent advantages from the use conventional pneumatics. The use and packaging pattern with electric counterparts needs to be followed to a certain level to gain recognition from industrial users. Design and support software tools needs to further developed. Well documented easily understood literature on servo pneumatic actuation needs to be produced, which at the same time address the practicality encountered by industrial users.

7. ACKNOWLEDGMENTS

The authors would like to acknowledge our sincere gratitude to UK EPSRC for providing the research funding (Grant No: GR/K 38663) on the project of High Speed Servo Pneumatic Actuator Systems

REFERENCES

[1] J Pu, R Harrison, P R Moore and R H Weston, "Design and Usage of Intelligent Servo Pneumatics as an Alternative to Electric Servo Technology", Machine Actuators and Control, Technical Seminar, I.Mech.E., London, 31 March 1993, pp53-59

[2] J Pu, P R Moore and C B Wong, "Commissioning Servo-Pneumatic Systems for Timing and Profile Control in High Speed Food Packaging Applications", Int. J. Prod. Res., vol. 33, No. 10, 1995, pp2907-2922

[3] P T Mo, "Modelling of Servo-controlled Pneumatic Drives - A Generalised Approach to Pneumatic Modelling and Applications in Servo-Drive Design", PhD thesis, Loughborough University, UK, 1989.

[4] D Scholz and R Schwenzer, "Pneumatic Proportional Valve with Integral Electronics - Festo MPYE 5-1/8" Aachen Fluid Power Colloquium, 17-19 March, RWTH Aachen, Proceedings, pp129-140

[5] R Bazany, "Choosing a Control Network", Control Engineering, July 1997.

[6] S K Yung and D W Clarke,"Local Sensor Validation", Measurement & Control, Vol. 22, June, 1989

[7] "Optimised Reference Manual", Baldor-Optimised Control Ltd.,

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Modelling Framework for Design and Analysis of Distributed Real-Time Control Implementations

O. Redell and M. Törngren

Mechatronics Lab, Department of Machine Design, Royal Institute of Technology
S-100 44 Stockholm, Sweden, email: {ola, martin}@damek.kth.se

The AIDA decentralization tool-set, intended for design and early implementation analysis of control systems, is currently being developed at the Mechatronics Lab, KTH (Royal Institute of Technology). Special focus of the tool-set is control systems that are implemented in distributed computer networks. The control applications pose hard real-time requirements on the implementation and the tool-set is designed to support pre-implementation analysis. This paper describes parts of the modelling framework that the tool-set will be based on. The models are used to specify the control application, put requirements on the implementation and to describe the results of an implementation. Also the distributed hardware and the non-application software is described by the models.

1. INTRODUCTION

The project *Automatic control in distributed applications* (AIDA), treats problems associated with the implementation of control applications in distributed computer systems. Earlier work at the Mechatronics Lab [1], and input from industrial contacts have made it natural to focus on one of the most recent areas of distributed computer systems; machine embedded control systems. The number of industrial applications involving such systems are increasing rapidly due to cost-, functionality-, and performance-benefits of distributed control implementations. In a machine embedded distributed control system, electronic hardware (microprocessors, ASICs, mechanical interfaces, etc.) and control software is distributed in the machine and connected to sensors and actuators. The nodes of the distributed computer system are connected through communication networks making the closing of control loops possible.

The main motivation of AIDA is the lack of research in the gap between control system design and analysis, and its implementation on a distributed hardware. Examples of this are the computer science oriented behaviour models developed for real-time scheduling which are not sufficient for sampled data control applications, [1], while control engineering models on the other hand, do not take implementation in distributed computer systems into account. There is none or at least very little support for control system implementation analysis in today's tools for Computer Aided Engineering (CAE). This holds for both the control engineering (CACE) and software engineering (CASE) tools (e.g. tools supporting rapid prototyping, Matlab/Simulink, MatrixX and Software through Pictures, SDT and Teamwork respectively). Implementation support is limited to code generation and download of binary code to specific target hardware, after which that specific implementation can be tested. This is what is referred to as rapid prototyping.

The lack of tool support for control system implementation is the motivation for the development of the AIDA decentralization tool-set.

The tool-set is to be used for design and early implementation analysis of control systems that are implemented in distributed computer architectures. The tool-set is planned to support both initial design, augmentation of already existing systems and analysis of the resulting implementations. The tool-set will not include the functionality of e.g. commonly used CACE tools such as Matlab/Simulink, but will rather form a real-time design and implementation complement to such tools.

The scope of the tool-set puts hard requirements on the models on which the tool-set will be based. The models must be capable of describing a wide range of control system implementations unambiguously. Such models need to include description of both the application design, the implemented application, the computer hardware and for some cases also a limited mechanical description.

The current status of the tool-set development is that a specification has been developed in close cooperation with industrial partners. This work has led to the proposal of a modelling framework on which the tool-set will be based. The modelling framework has been used for a case study to model a distributed control system implementation for a four legged vehicle.

In section 2, some related work will be described. Then the intended use of the tool-set is outlined in 3, after which a sub-set of the models that the tool-set is to be based on is described in 4. In section 5, the work is concluded and future work discussed.

2. RELATED WORK

State of the art CACE tools do support rapid prototyping for specific distributed computer systems. The tool set provided by the companies Mathworks and dSPACE enable control system design, analysis and simulation, code generation from the graphical block diagrams of Simulink, "transparent" implementation of generated code on DSPs, and real-time operator interaction and logging, [3]. Transparency here refers to the fact that the user manually specifies allocation of Simulink blocks to processors, the physical configuration of which is also described in Simulink. The required communication code is then included transparently and a real-time scheduler can manage the execution of multirate systems.

While such tools are extremely useful for control designers, there is still a large amount of functionality missing with respect to implementation analysis. Tools, such as those mentioned in section 1, do not allow exploration of implementation alternatives, for example with respect to resource structures and execution strategies, nor do they provide any guarantee that an implementation will work with respect to real-time behaviour.

Out of a number of related research tools we have found the three following to be highly interesting, [4,5,6]:

- The Development Framework, [4], incorporates some of the functions and a tool structure similar to that we envision. The control design is specified in Simulink and then transferred to a commercial CASE tool (Software through Pictures) which models can be analyzed and transformed by other tools (clustering, replication). The combination of CACE and CASE technology facilitates multi disciplinary project development and makes use of the strong points of the respective modelling and analysis features. Hardware modelling, execution and communication strategy considerations appear to be weak points in the Development Framework.
- The GRAPE toolset, [5], is developed for digital signal processing systems and has an interesting tool structure, including tools for allocation, scheduling and restructuring of the computer hardware. Only point to point architectures are considered.
- A very interesting effort about which there is not much material is the Parallel Scalable Design Toolset (PSDT) from Honeywell research. PSDT is a CASE tool specifically addressing implementation in distributed and parallel real-time systems, [6].

The amount of theoretical research work is vast including work on general purpose parallel processing, parallel and distributed real-time systems, and related work in control engineering. See e.g. [1] for an overview.

3. INTENDED USE OF THE TOOL-SET

The AIDA tool-set is to be used for both new designs and to add functionality to already implemented control systems. The tool-set will be a complement to existing CAE tools, adding possibilities to model and analyse e.g. timing and triggering of control functions when implemented on different hardware architectures.

In the task of performing an implementation of a control system on a potentially unspecified distributed computer hardware architecture lies a number of design choices. The interactions between the different design choices are considerable, having a great impact on the final system performance. Therefore the AIDA tool-set is intended to support iterative engineering, answering a number of "what if" questions put by the designer.

Some of many important issues a designer has to consider and that the AIDA tool-set is expected to support are:

- *Control application structuring*. How are the application functions structured? The application may be restructured to e.g. make better use of parallelism.
- *Partitioning*. How is the application partitioned into processes? Relates to e.g. allocation and scheduling.

- *Allocation and scheduling.* Allocation of processes to processors and scheduling of the processes strongly affect the system performance.
- *Selection and structuring of computer hardware.* Relates directly to cost and performance, adaptability etc.

Typical analysis results that the AIDA tool-set is to give are: estimates of feedback delays, delay variations, jitter in sampling periods, processor and communication link utilisations, qualitative comparison between different application hardware structures, bottleneck identification, etc. These results can be used to answer a number of different design questions. An example is, "With fixed hardware and application structures, what performance can be achieved by (i) changing the hardware components, (ii) changing policies for scheduling, clock synchronisation etc., (iii) relaxing some application requirements such as jitter and delay tolerances?".

4. THE MODELLING FRAMEWORK

The systems to be analysed and hence modelled are typically time-triggered, multirate, control applications with different modes of execution. Limited support for event-triggered functionality is however necessary. These control systems are to be implemented on distributed heterogeneous hardware with both serial and parallel communication links interconnecting the processing elements (i.e. processors, ASICs etc.). The processing elements can in their turn be placed on different locations in the mechanical structure.

In the AIDA tool-set, modelling is divided into three separate parts, *application*, *computer system* and *mechanics*. The application models describe data and control flows as well as timing requirements of the motion control system. Furthermore, the functions and the inter function communications that constitute the smallest building blocks of the application models are described in detail in terms of resource requirements. In the computer system models, the processes and their interconnections resulting from partitioning, are described together with the computer hardware. Operating system behaviour, scheduling policies etc. are also defined. Physical locations of the processing elements are given in the mechanics models. The AIDA modelling framework is more thoroughly described in [2].

A sub-set of the models will be briefly described in the coming sections. The description is based on a case study in which models of the control system of a four legged vehicle were derived. The vehicle is currently being built at the Mechatronics lab. The complete case study is described in [2].

In the case study, fixed vehicle mechanics is assumed as is the design of the control application. However, two choices of hardware architectures are investigated. In the case study, the control system of the vehicle is modelled and a coarse analysis of its implementation is performed.

4.1. Application models

The application is the control design as it is originally modelled in e.g. Simulink. In the AIDA application models, the control design is further specified with three basic types of diagrams: The macro mode transition diagram, data flow diagrams and timing and triggering diagrams.

The Macro Mode Transition Diagram (MMTD) describes the different high-level control modes of the system. An example is the take-off, cruising and landing of an aircraft. Each mode can be expected to use substantially different controllers that are best described separately.

The Data Flow Diagrams (DFD) describe the data-flows between the different control functions within each macro mode. The DFD is similar to the diagrams in common CACE tools, such as Simulink.

Figure 1 shows a part of a data flow diagram for the control system of the vehicle mentioned above. The torque control loop of one joint of the vehicle's four legs is shown. The diagram shows how functions S, TC and A are connected to perform a sample (S), compute (TC) and actuate (A) loop. The position samples include measures of two joint angles (on both sides of the flexible joint), used by TC to compute the joint torque.

Data flows to functions not shown in Figure 1 are indicated by dashed lines. Those data flows correspond to the position sample, (pos), sent to an observer function higher in the hierarchy and the reception of a torque reference signal, (torq), from a higher level controller.

The behaviour of the application in terms of timing, triggering and precedence is modelled in TTDs (Timing and Triggering Diagrams). The same functions are used as building blocks as in the DFDs, but the TTDs separate functions that are executed with different periods into activities. A TTD describes the

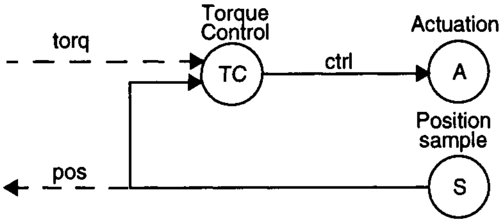


Figure 1. Data flow diagram for torque control loop.

control flow of the activities within each macro mode. Precedence relations, time and event triggers are shown and tolerances can be given for the control delays, jitter and synchronisation. This is useful to specify the requirements that the control design poses on the implementation.

With the example functions in Figure 1, a TTD can be specified. It is assumed that all functions in the torque control loop are specified to be executed with the same frequency (period) but that the actuate function should be triggered 0.5 ms after the sample function in order to achieve a constant control delay. Hence, the functions are collected into one activity, described by one single TTD. It is further assumed that there are requirements on the admissible jitter of the sample function, as well as on the variation in control delay. These requirements are shown in the TTD as specifications on the time trigger. A time trigger is defined as follows: `<Type: Source, Period: [Delay], Tolerance>`. The *Type* describes if it is a time, event or loop trigger (TT, ET or LOOP). The *source* indicates what the source of the trigger is (a clock for time triggers and an event for event triggers). *Period* gives the period of a time trigger and minimum period of an event trigger. *Delay* is an optional specification of the delay of the trigger compared to other triggers having the same source. Finally, the *tolerance* defines a tolerance of what variation is admissible from the specified triggering time.

Figure 2 shows the TTD for the torque control loop in Figure 1. The precedence relations between the functions are shown as well as the explicit triggers that both the sampler (S) and actuate (A) functions are given. It is specified that the actuation should be done 0.5 ms after the sampling. Both the sample and actuate functions are allowed to deviate with at most 100 μs from their specified start times.

The double rectangle defines a *rate interfacing function* which is used to describe asynchronous (non blocking) communication between different activities. The meaning is that the sample function sends data ('pos' in Figure 1) to an observer function, (Obs), that belongs to the parts of the control system that are not modelled here. The observer function should however be part of some other TTD in a complete system modelling. In a similar manner, a rate interfacing function (tagged 'TC') is used in some other TTD (which is not given here), to describe the communication associated with the data flow 'torq' in Figure 1. In that case, TC is the receiving function.

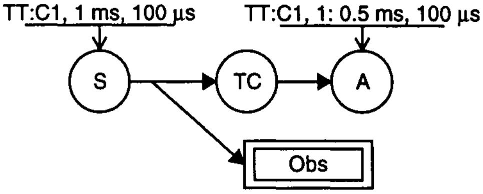


Figure 2. Timing and triggering diagram for the torque control activity.

A timing and triggering diagram can describe parallel paths as well as alternative paths, selected at runtime. A path that is not necessarily executed every time the activity is triggered is called a micro mode. Two complementary micro modes may e.g. be the use of a fast linear controller versus using a better but slower nonlinear one. It should be noted that the arbitration between different micro modes is not described at a low level with e.g. if-then-else constructs. Micro modes are included only to indicate possible execution paths in an activity, not to show how the micro modes are selected.

One additional timing and triggering diagram is used to describe the timing that results from an implementation. The Implementation TTD (I-TTD) is expected to be the result of some analysis tool within the AIDA tool-set. An I-TTD is similar to its corresponding original TTD but with the difference that extra delay functions have been included to show the effects of implementation. Furthermore, the triggers are specified with their achieved values on jitter, delays and period. The added delay functions include delays originating from e.g. scheduling effects, blocking and sequential execution of concurrent activities.

Figure 3 shows an I-TTD for the torque control activity in Figure 2. The delay functions D_1 and D_2 have been included and the values specifying the triggers have been exchanged for some real values obtained by analysis.

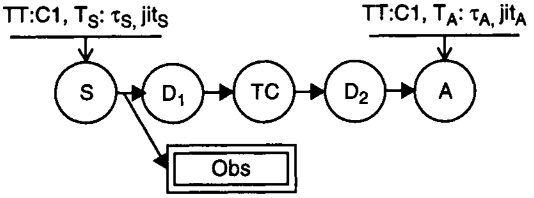


Figure 3. Implementation timing and triggering diagram for the torque control activity.

At the functional level of the application models, the functions and the data flows are described with worst case execution times, communication requirements etc. Hence, the functions are not described in detail, only their resource needs are specified.

4.2. Computer System

The computer system models include specification of the computer hardware, the processing elements and their inter connecting communication links. Also, the operating system is described in terms of scheduling policies, dispatching time, clock tick period etc.

When the application has been partitioned into processes, the connections (communications) between the processes are shown in a process structure diagram. The inter process communications (IPCs) are modelled with communication resources (message queues and transmitters) that specify e.g. blocking times of message queues and scheduling policies of the communication links. The communication models can later be used for analysis of e.g. communication latencies.

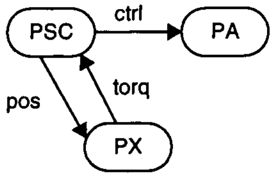


Figure 4. Process structure diagram for the three processes in the example.

Assume that the example activity shown in Figure 2 is partitioned into two processes—PSC and PA. The partitioning may correspond to e.g. a distributed implementation with the sample and computation functions at one processor, and the actuation at another. PSC contains the sample and computation functions (S and TC), and PA contains the actuation function (A). Also assume that another process, PX, corresponding to all other activities in the system, is defined. The corresponding process structure diagram is shown in Figure 4. Note that three IPCs are defined: pos, torq and ctrl. Each of the IPCs are data flows in Figure 1 that “cross process borders”. For simplicity, the IPCs are here given the same tags as their corresponding data flows in the DFD.

The timing of the processes is specified in a Process timing and triggering diagram (P-TTD) that looks like a regular TTD only that it specifies the timing and triggering of the processes and not of the functions. In Figure 5, a P-TTD is shown that defines the requirements of the process timing. The same types of triggers are used as in the regular TTD but no rate interfacing functions are included. Instead, the correspondence to rate interfacing functions, the inter process communications, are visible in the Process internal timing and triggering diagrams (Pi-TTD). A Pi-TTD defines the control flow internal to a process with the functions from the original TTD as building blocks. Also the IPCs are visible as blocks defining the reception and sending of a message. Furthermore, any functions that are allowed to execute in parallel according to the original TTDs, are shown in sequence. The blocks corresponding to the sending and the reception of a message are associated with communication resources belonging to some communication model. The communication resource blocks are shown as rectangles in the Pi-TTD.

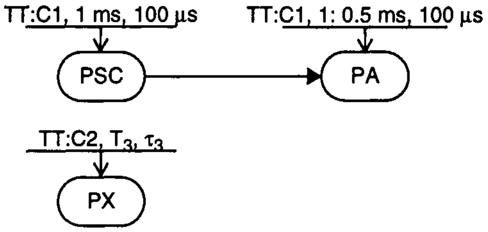


Figure 5. Process timing and triggering diagram for the three processes in the example.

In Figure 6, the Pi-TTDs for the processes PSC and PA are shown. The first of the diagrams shows the PSC process with visible blocks for interfacing the sender and receiver communication resources. Sending of a position message (pos) is shown as the block S_pos, reception of torque reference is tagged R_torq etc.

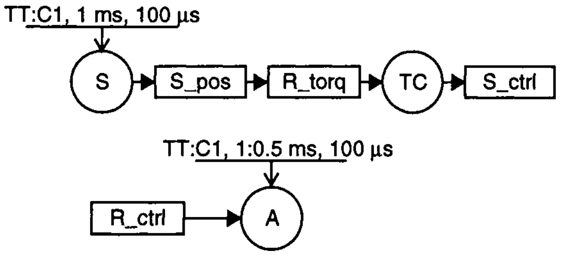


Figure 6. Process internal timing and triggering diagrams for process PSC and PA.

Note that all the diagrams shown in this example are specification diagrams used to specify a required behaviour rather than showing the results of an implementation. Just like in the TTD and I-TTD case however, the diagrams can be used with achieved values of the time triggers to show a result rather than a specification. In that case, triggers are assigned to the starting block of each process. Compare Figure 6 where the requirement specification is given for the timing of the actuate function.

4.3. Mechanics

The mechanical models are included in the modelling framework primarily to support optimization with respect to cabling costs, and to leave a possibility to later include other mechanics related optimization criteria.

The mechanical models describe locations in which a designer finds it reasonable to place processors or environment interfaces such as e.g. drive units. The locations are given cable distances as coordinates. A cable distance is the length of the cable that is needed to connect one location with another. A location is usually associated with a number of sensors and actuators. These are also described in the mechanical model.

5. CONCLUSIONS

The models described in this paper are just a subset of the models in the AIDA modelling framework. The parts that were not described handle among other things scheduling policies for the processors and inter processor communications via communication links and via operating system message queues etc. Also the hardware components and structure is described. The execution time analysis that is supported by the models is however coarse. Functions for more accurate execution time analysis is not planned to be a part of the AIDA tool-set even though it may be of interest to include such a tool in the future.

An early version of the modelling concept was tested in [1], while the more detailed case study is described in [2]. More case studies are needed to further verify that the models are capable of describing the types of systems the AIDA tool-set aims for.

Future work will include further refinements and more exact semantic definitions of the models. Functions for the AIDA tool-set are also to be developed primarily to support basic implementation analysis. At a later stage tools for design such as (semi-) automated allocation, scheduling and structuring are to be added.

6. ACKNOWLEDGEMENTS

This work was funded by Branschgruppen för Datorstyrd Mekanik, the Swedish National Board for Industrial and Technical Development (NUTEK) and the Volvo Research Foundation.

REFERENCES

[1] Törngren (1995), *Modelling and design of distributed real-time control applications*. Doctoral thesis, Department of Machine Design, KTH, TRITA-MMK 1995:7, ISSN1400-1179, ISRN KTH/MMK--95/7--SE.

[2] Redell, *Modelling of Distributed Real-Time Control Systems, An Approach for Design and Early Analysis*, Licentiate Thesis, Dep. of Machine Design, KTH, 1998, TRITA-MMK 1998:9, ISSN 1400-1179, ISRN KTH/MMK--98/9--SE

[3] Hanselmann (1995), *DSP in Control: The Total Development Environment*. Int. Conference on signal processing applications and technology. Oct. 24-26, 1995, Boston, MA, USA.

[4] Bass et al. (1994), *Automating the Development of Distributed Control Software*. IEEE Parallel & Distributed Tech., Winter 1994.

[5] Lauwereins et al. (1995), *Grape-II: A system-level prototyping environment for DSP applications*. IEEE Computer, Feb. 1995, pp. 35-43.

[6] Bhatt (1996), *A methodology and toolset for the design of parallel embedded systems*. In Proc. of the School on Embedded Systems, Organised by the European Educational Forum, Veldhoven, NL, Nov. 1996

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Robust Motion Control for Planar Laser Cutting Machine

Aleš Hace, Karel Jezernik, Boris Curk, and Martin Terbuc

Institute for Robotics, Faculty of Electrical Engineering and Computer Sciences, University of Maribor
Smetanova ul. 17, SI-2000 Maribor, Slovenia

A position control algorithm for a belt-driven servomechanism of a laser-cutting machine is described. High-accuracy position tracking control procedure for the system with inherent elasticity due to the low-cost belt-driven servomechanism is derived based on continuos sliding mode technique. The proposed robust position tracking control algorithm is used in an industrial application of a motion controller for the CNC laser-cutting machine. Experimental results are reported.

1. INTRODUCTION

Performances of movement transformation elements from motors to a machine tool play important role in a design of a motion controller. A mechanical construction of the machine and built-in drives define drive's performances. There are two major sources of uncertainties in the machine with DOFs in horizontal plane: friction and inertia. When belts are used, instead of assuming a rigid drive system, a non-rigid drive system with additional uncertainty of belt elasticity (stiffness) should be incorporated in the controller design.

In recent years, disturbance observer technique [1], [2] or disturbance estimation technique [3] are often used in a motion control of a mechanical system. High-speed motion response under the high accuracy positioning control in a rigid spindle driven servo drives is a task, where the mechanical vibrations arise. Therefore, a feedback control structure with the ability of vibration suppressing should be designed [1]. Even more mechanical vibrations arise when belts combined with high loads are used in drive system. Vibrations arise due to the belt-stretch oscillation [4]. In order to control belt-stretch and to stabilize belts oscillation, which consequently influence position-tracking error, a vibration-suppressing controller is neglectible. Therefore, new robust control algorithm based on [3] and improved with feedback structure to suppress arisen vibrations due to belt-driven motion system in 2DOF machine for laser cutting is derived.

In our paper control algorithms of the biaxial motion controller for the laser-cutting machine are represented. The paper is organized in following sections: the 2nd section briefly describes the machine and mechanical models of driven axes, the 3rd and the 4th sections represent control algorithms used in motion controller, and in the 5th section simulation and experimental result are reported, which follows with conclusions in the 6th section.

2. CONTROL OBJECT

The laser-cutting machine consists of a XY horizontal table and subsystems needed for the process of a laser cutting (Fig. 1). Both axes are driven by DC servomotors with permanent magnets. A rotational movement is reduced by gears and transformed to a translational movement of a laser head by belts. The laser head is placed on a bridge and represents dominant load of 20kg in X direction. Original construction solution results in very high load of 250 kg in the Y direction. Due to a low-cost belt-driven motion system combined with high loads additional disturbances arise. Thus, transformation of movement from the motor shaft to the machine

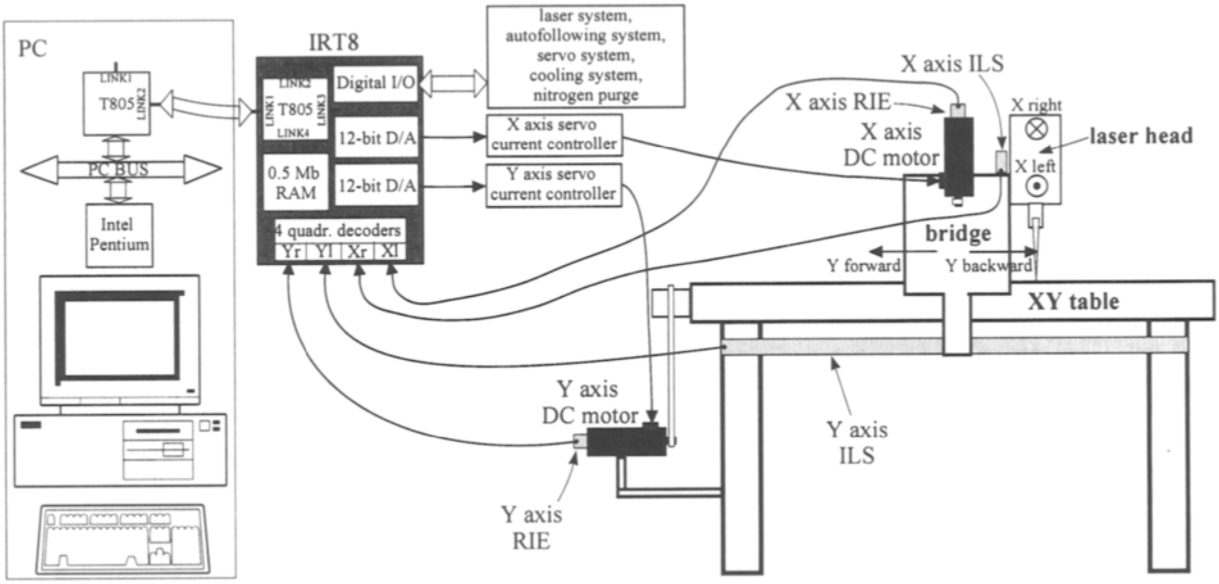

PC
LINK1
T805
LINK2
PC BUS
Intel Pentium
IRT8
LINK2
LINK1
T805
LINK3
LINK4
Digital I/O
0.5 Mb RAM
12-bit D/A
12-bit D/A
4 quadr. decoders
Yr
Yl
Xr
Xl
laser system, autofollowing system, servo system, cooling system, nitrogen purge
X axis servo current controller
Y axis servo current controller
X axis ILS
X axis RIE
X axis DC motor
X right
X left
laser head
bridge
Y forward
Y backward
XY table
Y axis DC motor
Y axis ILS
Y axis RIE

Fig.1. The machine and the controller hardware

tool (the laser cutting head) is uncertain. Arisen vibrations in the motion and inaccurate movement could be compensated only if position informations of motor shaft and laser cutting head are available simultaneously. Therefore, the position measuring system consists of rotary incremental encoders (RIE), which are placed on the motor axes, and incremental linear scales (ILS), which are placed on the mechanical construction of the machine. RIEs have a resolution of 8000 p/r and 40000 p/r for the X-axes and Y-axes, respectively. ILSs have the resolution of 1p/µm.

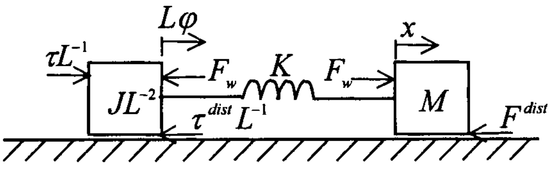

$L\varphi$
x
τL^{-1}
JL^{-2}
F_w
K
F_w
M
$\tau^{dist}L^{-1}$
F^{dist}

Fig. 2. Mechanical model of non-rigid single axis

The belt-driven servomechanism represents a distributed parameter system. Using a modal analysis, it could be modeled as a multi mass system. Motor inertia (J) and load mass (M) coupled with a spring (belt stiffness K), and friction on the motor side and load side could be dominant properties of

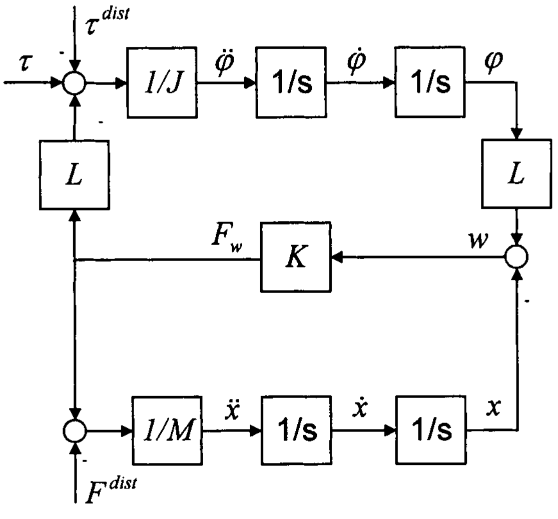

τ^{dist}
τ
$1/J$
$\ddot{\varphi}$
1/s
$\dot{\varphi}$
1/s
φ
L
L
F_w
K
w
$1/M$
$\ddot{x}$
1/s
$\dot{x}$
1/s
x
F^{dist}

Fig. 3. Scheme of the control plant model

physical model with concentrated parameters. Such a mechanical model of single belt-driven axis is featured by Fig. 2. Since the control objective is to control position of the laser beam (x) which serves as a tool, Fig. 2 shows the model with motor inertia transformed on the load side. Therefore, the motor angle position (φ), applied torque (τ), and disturbance torque including friction torque (τ^{dist}) are transformed on the load side, too. F_w denotes spring

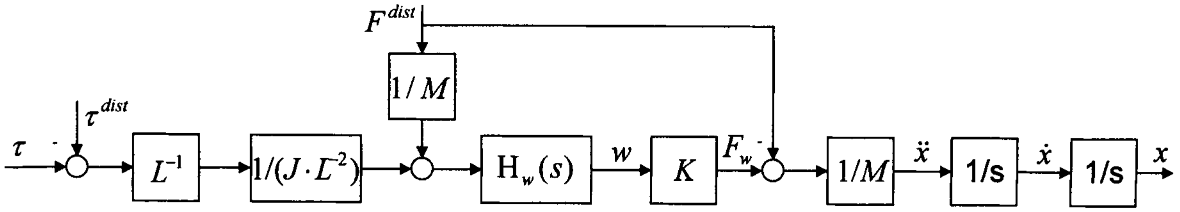


Fig. 4. Modified scheme of the control plant

force. F^{dist} denotes disturbance force in the motion system due to the nonlinear friction and unmodeled dynamics on the load side. L denotes transformation constant ($L = \partial x / \partial \varphi$).

Block scheme of the control plant derived from the mechanical model from Fig. 2 is featured by block scheme in Fig. 3, where w denotes the belt stretch. The model is featured by φ / x state-space. It involved coupled loops, which should be resolved in order to simplify drive dynamics representation. The simplification results in w/x state-space model. Modified model is represented by Fig. 4. The belt stretch w occur due to the excitation of applied torque τ and disturbances τ^{dist}, F^{dist}:

$$w = H_w(s)\cdot\left(L^{-1}\cdot\tau / JL^{-2}\right) + w^{dist}, \tag{1}$$

$$H_w(s) = \frac{1}{s^2 + K\cdot\left(1/JL^{-2} + 1/M\right)}, \tag{2}$$

$$w^{dist} = H_w(s)\cdot\left(F^{dist}/M - L^{-1}\cdot\tau^{dist}/JL^{-2}\right). \tag{3}$$

The load side dynamics is

$$M\cdot\ddot{x} = F_w - F_{total}^{dist}, \tag{4}$$

where F_{total}^{dist} denotes total disturbances which influence load side dynamics

$$F_{total}^{dist} = F^{dist} + K\cdot w^{dist}. \tag{5}$$

3. POSITION CONTROL ALGORITHM

The motion controller for the laser-cutting machine has to be built in order to executes circular and linear movement, which are classified as CP movements. Therefore, the motion controller has to be designed with a position trajectory tracking control algorithm. In [3] robust position trajectory tracking control algorithm is represented. It is based on the perturbation estimator, which assures a continous sliding mode regime for a closed-loop system. Using perturbation estimator unmodeled uncertainties (e.g. friction...) are effectively compensated. The control algorithm is derived for a rigid mechanism

$$M\cdot\ddot{x} = F - F^{dist}, \tag{6}$$

where F and F^{dist} denote applied control force and disturbance force due to the perturbations present in then control system, respectively. The control law [3] assures an attractiveness of the sliding mode manifold,

$$S = \left\{x, \dot{x}, \ddot{x} : \left\|\sigma(t, x, \dot{x}, \ddot{x})\right\| < \delta\right\}, \tag{7}$$

in a finite reaching time, where a time-varying function σ is chosen based on a second order position tracking control objective

$$\sigma := \left(\ddot{x}^d(t) - \ddot{x}\right) + K_v\cdot\left(\dot{x}^d(t) - \dot{x}\right) + K_p\cdot\left(x^d(t) - x\right), \tag{8}$$

and δ denotes arbitrarily chosen vicinity of the sliding manifold [5]. Positive constants K_v and K_p shape the desired dynamic position error tracking behavior.

Motion in the vicinity of the sliding mode is refereed as the sliding mode motion based on definitions in [6], which ensures robustness against to the perturbations. The finite reaching time into a sliding mode regime and robustness are achieved using perturbation estimator based on the predefined very simple control model

$$M\cdot\ddot{x}^c = F^c, \tag{9}$$

where index $(.)^c$ denotes model variables. The model is used to calculate position tracking control law. The sliding mode regime could be achieved, when the control plant behaves as the model. In order to assure model dynamics behavior (9), which could be observed by the acceleration error $\ddot{x}^c - \ddot{x}$, the applied control force is decoupled in two parts

$$F = F^c + \hat{F}^{dist}. \tag{10}$$

The first part (F^c) is defined based on the sliding mode manifold definition (7), (8) and the desired model behavior (9)

$$F^c := M \cdot \left(\ddot{x}^d + K_v \cdot \left(\dot{x}^d - \dot{x}\right) + K_p \cdot \left(x^d - x\right)\right). \tag{11}$$

Note, that from (9) and (11) follows ideal model acceleration control

$$\ddot{x}^c = \ddot{x}^d(t) + K_v \cdot \left(\dot{x}^d(t) - \dot{x}\right) + K_p \cdot \left(x^d(t) - x\right). \tag{12}$$

The second one ($\hat{F}^{dist}$) denotes perturbation estimation. Since the perturbation estimation error $F^{dist} - \hat{F}^{dist}$ results in acceleration error

$$\sigma = \left(F^{dist} - \hat{F}^{dist}\right)/M = \ddot{x}^c - \ddot{x}, \tag{13}$$

which influences position tracking error dynamics, asymptotic stable estimator is neglectible in order to assure robust sliding mode regime. From (13) follows

$$M \cdot \sigma + \hat{F}^{dist} = F^{dist}. \tag{14}$$

If $\hat{F}^{dist}$ is calculated as

$$\hat{F}^{dist} := B_{est} \cdot \int \sigma dt = B_{est} \cdot \left(\dot{x}^c - \dot{x}\right), \tag{15}$$

where $\dot{x}^c = \int \ddot{x}^c dt$, than follows asymptotic stable dynamics of disturbance estimation

$$M \cdot \dot{\sigma} + B_{est} \cdot \sigma = \dot{F}^{dist}. \tag{16}$$

If the condition on disturbance properties $\left\|\dot{F}^{dist}\right\| < \xi$ is fulfilled (ξ is positive constant), than σ converges to the arbitrarily chosen sliding mode vicinity $\delta \leq \xi / B_{est}$ in the finite reaching time [5]. Additionally, in the sliding mode regime the upper bound on the position tracking error is $|e(t)| \leq \delta / K_p$, and the error dynamics is predefined with

$$\ddot{e} + K_v \cdot \dot{e} + K_p \cdot e = \sigma \to \delta. \tag{17}$$

4. VIBRATION CONTROLLER

The control law defined with (10), (11), and (15) guarantees arbitrarily chosen asymptotic stable position tracking error dynamics. Unfortunately, the control plant as it is featured in Fig. 2 can not be treated as a rigid mechanism. Attractiveness of the sliding manifold (7) can not be guaranteed due to the oscillations, which occur due to the non-rigid servodrive. Thus, the control law must be upgraded with an algorithm, which is capable of suppressing the vibrations.

The force F_w (Fig. 4), which is applied to the load side, has to be modulate in order to achieve effective compensation of disturbances F_{total}^{dist} (4) and to control the nominal motion dynamics. Unfortunately, the force F_w does not represent input in the system. Input in the system is the applied torque τ on the motor shaft, which is transferred to F_w via transfer function $H_w(s)$ (2). Since $H_w(s)$ includes nonstable poles, the control algorithm (10), (11), (15) could not be employed in order to effectively ensure fast motion dynamics of the load side.

The idea for the design of feedback control structure is to make the non-rigid mechanism coupled to behave as a rigid mechanism. Consequently, the robust sliding mode control algorithm could be used to control position of the load side. It is obvious, that the task is feasible with an effective control of the belt stretch w. In this paper the approach of pole placement technique by state-space (w, $\dot{w}$) feedback and feedforward gain is used to assure fast and damped stable response via transfer function $H_w(s)$. Let us introduce proposed control algorithm, which assures suppressing vibrations. It is defined with equations (18)-(23).

$$\tau = L \cdot \left(J \cdot L^{-2} \cdot \ddot{w}^c\right), \tag{18}$$

$$\ddot{w}^c = K_w^{ss} \cdot w^d - \left(K_{wv} \cdot \dot{w} + K_{wp} \cdot w\right), \tag{19}$$

$$w^d = F / K, \tag{20}$$

where w^d and $\ddot{w}^c$ are the desired belt stretch and calculated belt stretch acceleration, respectively. K_w^{ss}, K_{wv}, and K_{wp} are state-space controller parameters, which are calculated as

$$K_w^{ss} = \omega_w^2, \tag{21}$$

$$K_{wv} = 2 \cdot D_w \cdot \omega_w, \tag{22}$$

$$K_{wp} = \omega_w^2 - K \cdot \left(1 / JL^{-2} + 1/M\right), \tag{23}$$

where ω_w and D_w denote desired eigen frequency and desired damping, respectively. Resulted belt-stretch dynamics is arbitrarily chosen:

$$\ddot{w} + 2 \cdot D_w \cdot \omega_w \cdot \dot{w} + \omega_w^2 \cdot w = \omega_w^2 \cdot w^d + \\ + \left(F^{dist} / M - L^{-1} \cdot \tau^{dist} / JL^{-2}\right) \tag{23}$$

The control algorithm introduced in (18)-(23) could be reduced in a calculation effort. Considering mass JL^{-2} as a gain in state-space controller (18) parameters could be rearranged. Additionally, resulted K_w^{ss} and K could be included in the model mass and perturbation estimator gain. Model mass M is affected and the force transfer dynamics is considered ($F \to F_w$). Therefore, the model from (9) is modified with effective mass present in the closed loop system

$$\hat{M} = M \cdot JL^{-2} / K \cdot \omega_w^2, \tag{25}$$

and estimator gain B_{est} is redesigned regarding $\hat{M}$.

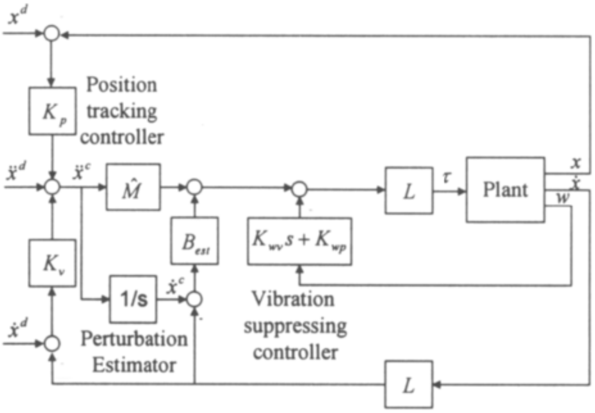


Fig. 5. Scheme of pos. tracking control algorithm

Resulted control scheme is featured in Fig. 5. It is transparently, that the controller consists of three parts: internal vibration suppressing state-space controller, the perturbation estimator that guarantees robustness, and position tracking part of the controller. The state-space controller parameters are calculated based on estimated values of plant parameters. Therefore, parameter model uncertainties are inherently included in the control algorithm. Due to the robust perturbation estimator, robustness of the overall control scheme is preserved.

5. EXPERIMENTAL RESULTS

Experimental tests were performed on the control object described in section II. The machine tip has to track the contour of an octagon in XY plane. At each corner the machine is stopped with prescribed position tolerance. The $\sin^2$ acceleration profile was used to generate reference position, velocity, and acceleration. Figures 6, 7, and 9,10 show position tracking results including position reference (on the top), position tracking error (in the middle), and applied reference voltage (on the bottom). Figures 8, 11 show octagons in the XY plane, where dashed line represents the reference contour and the solid one represents actual contour.

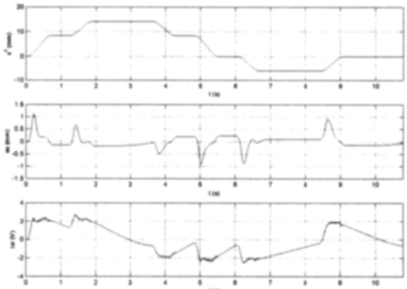


Fig. 6: X-axis pos. tracking - without vib. controller

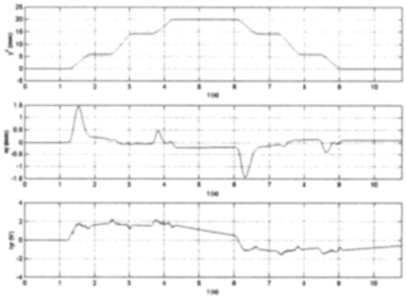


Fig. 7: Y-axis pos. tracking - without vib. controller

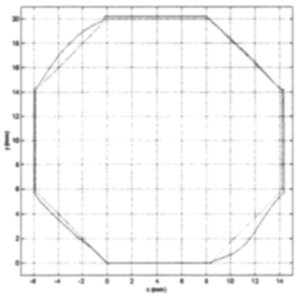


Fig. 8: Octagon in XY plane - without vib. controller

When vibration controller is not used a large tracking error (Fig. 6, 7) and consequently a large contour deformation (Fig. 8) could be observed. The error occurs due to the elasticity present in the system and the friction phenomena. Slow perturbation estimator dynamics can not effectively compensates disturbances. If vibration controller is employed, the servo drive behaves as a rigid one and fast estimator dynamics could be set up. Position error with a small magnitude occurs at velocity reversals due to the discontinuous friction changes, which is compensated very fast (Fig. 9, 10). Consequently, the contour deformation in XY plane practically disappears (Fig. 11).

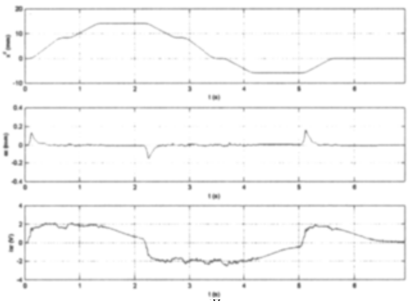


Fig. 9: X-axis pos. tracking - with vib. controller

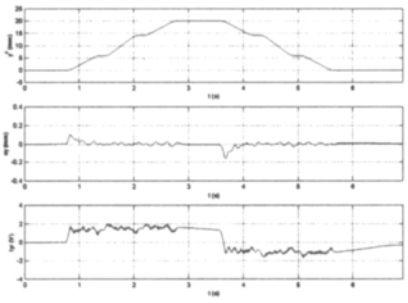


Fig. 10: Y-axis pos. tracking - with vib. controller

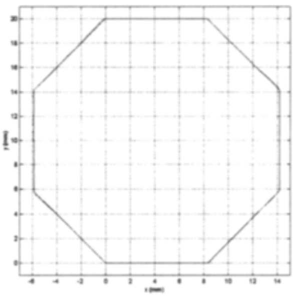


Fig. 11: Octagon in XY plane - with vib. controller

6. CONCLUSIONS

It is shown in the paper, that perturbation estimator designed to control rigid mechanism is effectively upgraded with feedback structure to suppress vibration in motion control of belt-driven servo-mechanism. Ability of compensating disturbances with fast estimator dynamics is achieved with pole-placement design of vibration controller. Unfortunately, discontinuous changes of static friction can not be total compensated immediately with a linear compensator. Position tracking errors, which occur at velocity reversals, are nevertheless compensate fast with very simple perturbation estimator control structure.

7. REFERENCES

1. K. Ohnishi, et al., Motion Control for Advanced Mechatronics, IEEE/ASME Trans. on Mechatronics, Vol. 1, No. 1, March 1996, pp. 56-67.
2. H. Seong, M. Tomizuka, Robust Motion Controller Design for High-Accuracy Positioning Systems, IEEE Trans. on Industrial Electronics, Vol. 43, No. 1, February 1996, pp. 48-55.
3. K. Jezernik, et al., Observer-based Sliding Mode Control of a Robotic Manipulator, Robotica, Vol. 12, part 5, Sept.-Oct.1994, pp. 443-448.
4. G. Hirmann, A. K. Belyaev, Stabilitätsverhalten eines schnellaufenden Synchronriemens, Antriebstechnik 36, Nr. 6,1997, pp. 64-66.
5. A. Hace, et al., Robust Sliding Mode Based Impedance Control, in Proceedings of the 1997 IEEE International Conference on Intelligent Engineering Systems, 97TH8224, pp. 77-82.
6. V. I. Utkin, Sliding Mode Control Design and Chattering Problem, Electrotechnical Review, Vol. 60, No. 2-3, Ljubljana, 1993, pp. 75-81.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Controlling a Shape Memory Alloy Drived Robot Gripper by Fuzzy Logic

István Mihálcz[a], Péter Baranyi[b], Zsolt Balogh[c], Péter Korondi[d], László Valenta[e] and Attila Halmai[f]

[a,e,f]Technical University of Budapest, Department of Precision Mechanics & Optics,
Egri J. u. 1, Budapest, H-1111, Hungary, e-mail: mihalcz@fot.bme.hu

[b,c,d]Technical University of Budapest, Department of Automation,
Budafoki u 8, Budapest, H-1111, Hungary, e-mail: baranyi@elktro.get.bme.hu

Using electrically actuated shape memory alloy (SMA) as an “artificial muscle” in robot mechanisms efficiently increases the degree of freedom offering much more flexibility than widely adopted grippers, that are mostly driven by pneumatic or electric actuators. In many SMA applications the evaluation of strain is difficult. This parameter depends on the size and geometry of the SMA wire, on the heat exchange conditions with the surrounding medium, on the type of alloy, and on the pulse with modulation. However applying SMA actuators leads to a very difficult control problem as the feature of SMA is strongly non-linear and very different for each, more over it is not well known yet. Consequently, the use of conventional control theories are very difficult in this case. In order to eliminate this problem we propose a control system based on fuzzy logic algorithm to learn the special feature of the gripper and then to control it. The experience introduced in this paper is a gripper constructed with SMA controlled by fuzzy algorithm and with a camera for learning feedback. The effectiveness of the system will be shown. This work is a preliminary research an one finger for our future work to control whole robot hand.

1. INTRODUCTION

Smart materials have received increasing attention in recent years for their great potential to revitalise and revolutionise engineering application and design. Nitinol is a nickel-titanium alloy discovered for its shape memory effect by the Naval Ordnance Laboratory in 1960's[4]. The shape memory effect has led to applications in the aerospace industry for pipe coupling, fastening, electric connection, energy storage and thermal and electric actuators. Nitinol has recently gained its popularity in medical devices for its biocompatibility, fatigue resistance, and body-temperature superelasticity.

Using SMA in robot mechanisms efficiently increases the degree of freedom offering much more flexibility than widely adopted grippers, that are mostly driven by pneumatic or electric actuators [1,2,3,4,5]. In this paper a robot gripper with SMA actuators is described. The feature of the gripper cannot be defined in advance as the construction is very difficult in the sens of controlling (high degree of freedom and non-linear SMA actuators [1,2,3]) and the feature of each SMA actuator is different, more over it is not well-known yet. Consequently, the use of conventional control theories are very difficult in this case. Fuzzy control generally has been used in many robot control applications [6,7,8]. This is the reason why we used a control system based on fuzzy logic algorithm to learn the special feature of the constructed gripper and then to control it. Our learning phrase follows the first steps of human moving. First we randomly set control values on the actuators and we then detect the moving of the gripper by a camera. Then using these information we tune a fuzzy controller.

Constructing a fuzzy system to emulate an input/output relation has been the topic of numerous works. Depending on the implicit or explicit nature of the relation, and the manner the input/output data becomes available, these works can be grouped into areas of fuzzy identification, fuzzy learning and fuzzy approximation.

All the same, however, constructing a fuzzy system involves the determination of the membership functions, the rule consequent sets and the inference paradigm, and these three parts are normally independent on each other, making the problem difficult. The advantage of applying fuzzy logic technique lies in their ability to imitate and implement the actions of expert operator(s) without the need of accurate mathematical models. The disadvantage, however, is that there is no standardised frame work regarding the design, optimality, reducibility, and definition of the concept. These algorithms, either generated by expert operators, or by some learning or identification schemes, may contains redundant, weakly-contributing, or unnecessary components. Moreover, to achieve a good approximation, some approaches may overly estimate the number of fuzzy rules, thereby resulting in a large fuzzy rule base causing problems in computational time complexity. The singular value based (SVD) fuzzy rule base reduction published by Yam [9,10] can actually be used to compress the information included in the rule base and the shapes of fuzzy sets.

In our application we use a back-propagation learning algorithm to a fuzzy rule base, that has triangular shaped antecedent and singleton consequent sets. In order to generate the control value product-sum-gravity inference is used. We use four actuators in almost symmetrical spatial position. This implies that the controllers for each actuator have similar, but not the same rules. In this case SVD reduction method can be used to reduce the size of the storage space required for the rules.

The organisation of the paper is as follow: Section 2 describes the SMA equations, Section 3 briefly introduces the Fuzzy rule base, Section 4 introduces the single value based Fuzzy approximations. Section 5 describes the experimental set up and presents results, and Section 6 draws conclusions from the results.

2. THE FEATURE OF SMA

In our application we use superelastic Nitinol. Unlike most other alloys, superelastic Nitinol exhibits an elastic stress plateau before permanent plastic deformation. Although the elastic deformation is strongly non-linear, the deformation is almost completely recoverable and can tolerate a large strain (5-8%) [3].

The relationship between the temperature-driven and stress-driven phase transformation can be described by the classic Clausius-Clapeyron equation [3]:

$$ds/dM_s = -DH/Te$$

where s is the internal stress generated by an external force, M_s the martensite start temperature, $D\,H$ the enthalpy of the phase transformation, T the temperature, and e the transformation strain. Indeed, this relationship has been verified in various shape memory alloys.

Tanaka determined the basic governing equation in 1982 [3], and this law was later modified by Liang and Rogers in 1990 [1], and then by Brinson in 1993 [2]:

$$\begin{aligned}\sigma-\sigma_0 &= E(\xi)\cdot\varepsilon-E(\xi_0)\cdot\varepsilon_0+\\ &+\Theta\cdot(T-T_0)+\Omega(\xi)\cdot\xi_S-\Omega(\xi_0)\cdot\xi_{S0}\end{aligned} \tag{1}$$

Here σ is the stress in the wire, ε the strain, T the temperature, E the Young's modulus of the material, Θ the thermal coefficient of expansion, ξ the martensitic fraction in material, Ω the transformation tensor. The '0' subscripts denote the initial conditions. Because exist two type of martensitic transformation, induced by temperature and by mechanical tension, we can distinguishes twinned(ξ_T) and detwinned (ξ_S) martensite in the material [2]: $\xi = \xi_S + \xi_T$

The transformation tensor must be expressed as (Liang and Rogers, 1990; Brinson, 1993) [2]:

$$\Omega(\xi)=-\varepsilon_L\cdot E(\xi)$$

where ε_L is the maximum residual strain for the given SMA material. It can rewritten the general equation:

$$\begin{aligned}\sigma &= E(\xi)\cdot(\varepsilon-\varepsilon_L\cdot\xi_S)+\Theta\cdot(T-T_0)+K_0\\ K_0 &= \sigma_0-E(\xi_0)\cdot(\varepsilon_0-\varepsilon_L\cdot\xi_{S0})\end{aligned} \tag{2}$$

where the K_0 represents the initial conditions.

Consequently controlling SMA actuators leads to a difficult non-linear control problem.

3. CHARACTERIZATION OF THE FUZZY RULE BASE

The rule bases have two inputs and one output.

Characterisation of fuzzy sets

In order to save computational effort we use triangular fuzzy sets as $\{A_{1,1}\quad \mathrm{K}\quad A_{1,n_1}\}$ and $\{A_{2,1}\quad \mathrm{K}\quad A_{2,n_2}\}$, where n_1 and n_2 is the number of antecedent sets on input universe $X_1 = [0,1]$ and $X_2 = [0,1]$, respectively, so that (p=1,2)

$$\text{Core}\{A_{p,k}\} = a_{p,k},$$
$$\text{support}\{A_{p,k}\} = [a_{p,k-1}, a_{p,k+1}],$$
$$\text{support}\{A_{p,1}\} = [a_{p,1}, a_{p,2}],$$
$$\text{support}\{A_{p,n_p}\} = [a_{p,n_p-1}, a_{p,n_p}].$$

The consequent fuzzy sets $B_{i,j}$ are singleton sets as: $\mu_{B_{i,j}}(y) = \delta(y_{i,j})$; $y \in Y$. For the observation A^*_p we use singleton set as $\text{core}\{A^*_p\} = x_p$ where x_p is the input value on X_p.

Characterisation of the rules

$$\text{If } A_{1,i} \text{ and } A_{2,j} \text{ then } B_{i,j} \Rightarrow \delta(y_{i,j})$$

Characterisation of inference

We use fuzzy inference based upon product-sum-gravity [9].

Product: Each rule gives a contribution set of the corresponding consequences $B_{i,j}$ multiplied by the input membership degree. So, if antecedent $A_{p,k}$ has membership degree $\mu_{A_{p,k}}(x_p)$ then this rule yields a contribution set $B_{i,j}$ in the form

$$\mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)\delta(y_{i,j}).$$

Sum: All contribution sets are summed to produce the fuzzy set conclusion $\tilde{B}$ as

$$\tilde{\mu}_B(y) = \sum_i \sum_j \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)B_{i,j}$$

which might not be directly interpretable as a fuzzy set, as function $\tilde{\mu}_B(y)$ may be larger than 1. In this case it should be normalised from the theoretical viewpoint to obtain a real fuzzy set.

Gravity: We apply centre of gravity defuzzification to obtain output value y. Thus the output values are calculated as:

$$y = \frac{\sum_i \sum_j \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)y_{i,j}}{\sum_i \sum_j \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)} \tag{3}$$

Considering the characterisation of the antecedent sets the membership degrees of the antecedent sets at any value within the universe sum to 1. Thus

$$\sum_i \sum_j \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2) = 1$$

that means this rule base performs a piece-wise linear approximation as from (3).

The learning method does not tune all sets, but only the position of the consequent sets. Let $\underline{\mathbf{m}}$ be a vector that contains the membership degrees as:

$$\underline{\mathbf{m}} = [m_1 \ldots m_m \ldots m_M],$$

where $m_m = \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)$. $\underline{\mathbf{B}} = [y_m]$ contains the core of B_m, where $m = (i-1)n_1 + j$.

The steps of the algorithm are:

1) $\underline{\tilde{y}} = \underline{\mathbf{B}}^{\mathrm{T}}\underline{\mathbf{m}}$,

2) $\underline{\delta} = \underline{y} - \underline{\tilde{y}}$,

3) $\Delta\underline{\mathbf{B}}^{\mathrm{T}} = \eta\underline{\delta}^{\mathrm{T}}\underline{\mathbf{m}}$,

4. SINGULAR VALUE-BASED FUZZY APPROXIMATION

This section briefly summarises the Fuzzy Singular Value Decomposition (SVD) method of [9,10,11,12] for the fuzzy approximation of continuous function on compact domain. The approach calls for sampling the original function over a set of rectangular grid points and applying singular value decomposition to determine the shapes of fuzzy sets. Let $y_{i,j}$ be the sampled value of two variables function over the given grid point $g_{i,j}$, with i=1..n_1, j=1..n_2, and n_1 and n_2 are the respective number of grid lines defined by elements $x_{1,i} \in X_1$ and $x_{2,j} \in X_2$ on the input universe X_1 and X_2. A linear Fuzzy interpolation can be obtained between the sampled $y_{i,j}$ by a fuzzy rule base FL, using singleton observation x_1 and x_2 as follows:

$$f^{FL}(x_1, x_2) = \frac{\sum_i^{n_1} \sum_j^{n_2} \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)y_{i,j}}{\sum_i^{n_1} \sum_j^{n_2} \mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)} \tag{4}$$

where $\mu_{A_{k,i}}(x)$, $x \in X_k$ is the antecedent membership function of the i-th set $A_{k,i}$ defined on input universe X_k, and considering the shapes of the sets $A_{k,i}$ the following condition has to be satisfied:

$$\sum_i^{n_1} \mu_{A_{1,i}}(x_1) = 1 \text{ and } \sum_j^{n_2} \mu_{A_{2,j}}(x_2) = 1 \tag{5}$$

then: $$\sum_{i}^{n_1}\sum_{j}^{n_2}\mu_{A_{1,i}}(x_1)\mu_{A_{2,j}}(x_2)=1 \qquad (6)$$

where $x_1 \in X_1$ and $x_2 \in X_2$. The linear Fuzzy SVD method can then be applied to yield a reduced rule base FL^r. Let matrix **B** contains the element $y_{i,j}$. Applying RSVD, one obtains **B** = **A'**$_1$ **B'** **A'**$_2^T$ where **A'**$_1$ and **A'**$_2$ correspond to the antecedent sets and **B'** the consequent sets. The supposed antecedent sets **A'**$_1$ and **A'**$_2$, however, may not satisfy the condition (5). One hence applies SN to determine **B''**, **A''**$_1$ and **A''**$_2$ which satisfy (5) and (6). It is possible, however, to still have negative elements in **A''**$_1$ and **A''**$_2$. One hence further applies NN to determine **B'''**, **A'''**$_1$ and **A'''**$_2$ which correspond to positive membership degrees and satisfying (5) and (6) as well. The resulting rule base **B^r=B'''**, **A$^r_1$=A'''**$_1$ and **A$^r_2$=A'''**$_2$, where the number of rules is $c_1 \le n_1$ by $c_2 \le n_2$ as:

$$\mathbf{A}^r{}_i=\begin{bmatrix} A^r{}_{i,1}(x_{i,1}) & \cdots & A^r{}_{i,c_i}(x_{i,1}) \\ \vdots & \ddots & \vdots \\ A^r{}_{i,1}(x_{i,n_i}) & \cdots & A^r{}_{i,c_i}(x_{i,n_i}) \end{bmatrix};$$

$$\mathbf{B}^r=\begin{bmatrix} y^r{}_{1,1} & \cdots & y^r{}_{1,c_2} \\ \vdots & \ddots & \vdots \\ y^r{}_{c_1,1} & \cdots & y^r{}_{c_1,c_2} \end{bmatrix}$$

where $A^r{}_{i,k}(x_{i,n})$ is the value of the k-th membership function $A_{i,k}$ at the input point $x_{i,n}$.
For simplicity, the work of [9] adopts continuos antecedent sets which linearly connect the membership values $A_{i,k}(x_{i,n})$. The output value of the rule base FL^r is hence represented as:

$$f^{FL^r}(x_1,x_2)=\frac{\sum_{i}^{c_1}\sum_{j}^{c_2}\mu_{A^r_{1,i}}(x_1)\mu_{A^r_{2,j}}(x_2)y^r_{i,j}}{\sum_{i}^{c_1}\sum_{j}^{c_2}\mu_{A^r_{1,i}}(x_1)\mu_{A^r_{2,j}}(x_2)} \quad \text{with:}$$

$$\sum_{i}^{c_1}\sum_{j}^{c_2}\mu_{A^r_{1,i}}(x_1)\mu_{A^r_{2,j}}(x_2)=1 \text{ for } x_1 \in X_1,\ x_2 \in X_2.$$

If the discarded singular values δ_i of **B***$_d$ are not 0 then the error between sampled values and the output values of the rule base FL^r can be expressed as: $E \le \sum_i \delta_i$.

The linear Fuzzy SVD method can be viewed as rule base reduction method [10,11,12,13]. We introduce the use of SVD method for the rule base of one controller described in section 3.

This simple rule base linearly interpolates between the given grid points, and the number of rules is equal to the number of grid points (n_1 by n_2). The Fuzzy SVD method hence yields the linear interpolation FL^r as reduced rule base from FL using less antecedent sets depending on the samples of the original function. In extreme case when the information can not be compressed the Fuzzy SVD method approaches the same rule base as FL. It can be said from this viewpoint that the Fuzzy SVD method is a kind of fuzzy rule base reduction method which reduces the number of rules by reducing the number of antecedent sets. Extension to a general number of input is included in [9,10,11,12,13]. Applications of Fuzzy SVD to fuzzy rule interpolation can be found in [13]. Specifically, the work [12] studies application to various cases where the consequent sets are described by multiple characteristic points.

5. EXPERIENCE

The finger has four SMA's (figure 1.). The joint is a flexible silicon rubber, both of the two knuckles are vertical if the all SMA "muscles" are relaxed.

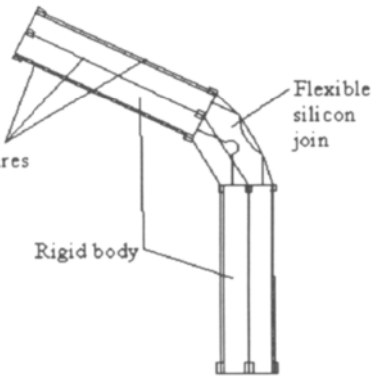


Figure 1.

The finger is moved by four (two pairs) symmetrically placed SMA wires. Each SMA is controlled one by one controller.
We develop a learning phrase as follows: we randomly drive the actuators so that we do not drive all, but always two actuators that do not move the gripper opposite direction.

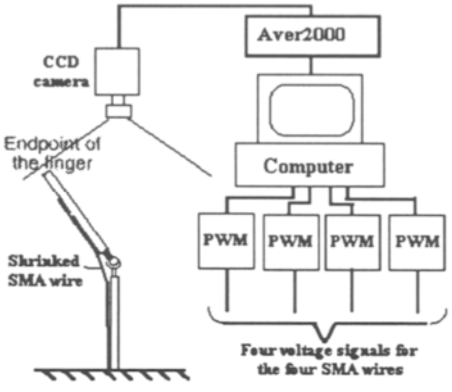


Figure 2.

(If e.g. two randomly driven actuators move the gripper into opposite direction it might destroy the construction.). Although, the endpoint of the finger is moving alone a spherical surface, the camera detects it as a planar movement (figure 2.).

Having the information, namely, the position of the gripper and the value the SMA driven by, we tune the controllers. For example position A of the gripper can be reached from B driving SMA a, b and from C driving SMA c, d etc. (see figure 3).

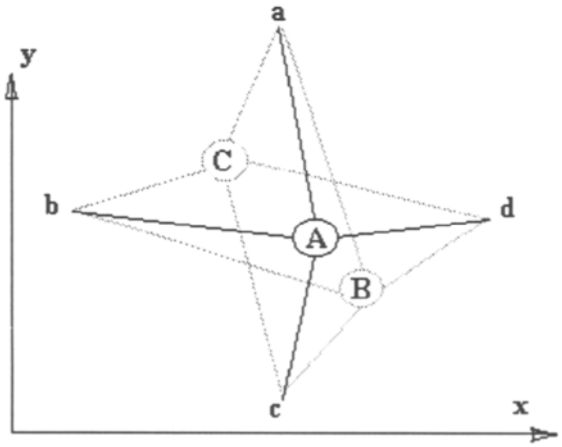


Figure 3.

So, each controller is tuned for position A in different cases. Consequently, if position A is required, after the learning phrase, all actuators will be driven, that means the gripper is uprighted by the four actuators.

The controllers have two inputs that define the position of the gripper and one output to drive SMA.

The fixing of SMA actuators are not symmetrical as the gripper is constructed in a simple way, however the use of the tuned controllers eliminates this problem.

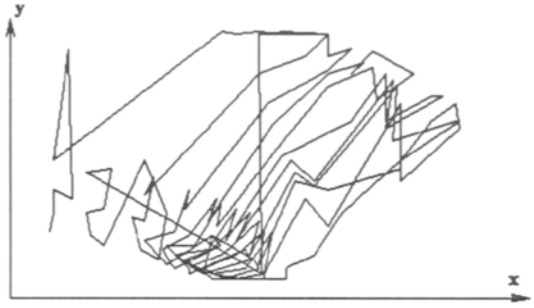


Figure 4.

For the experiment we used only two wires. Two PWM (Pulse Width Modulation) generators, commanded by a PC parallel port, powered these wires. Both outputs of PWM generators were coupled to current limiter circuits to protect the SMA-wires overheating [14]. The pulse width was randomly modified between 45% - 75%, as the pulse width under 45% has no effect and detected the moving of the end of robot gripper.

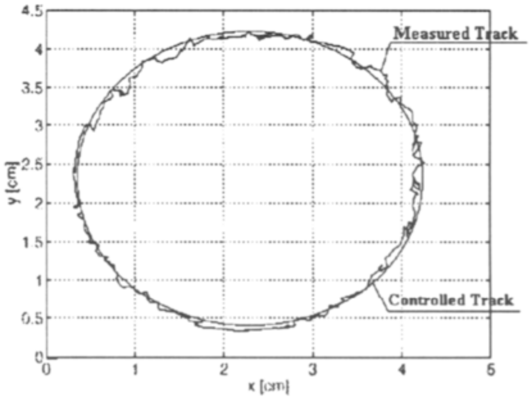


Figure 5.

Based on these information (see figure 4.) as learning patterns, namely the pulse width and position of the gripper, we tuned a fuzzy rule base.

In figure 4. the path is represented in xy plane (y=0).

After the learning phase we used the fuzzy rule base as a controller to the finger. The finger was moved alone a circular path using the reduced Fuzzy rule-base, the measurement results is shown in figure 5. After learning process each and every pont into the working space was catch with an accuracy of 1-2 mm.

6. CONCLUSION

In this paper a robot gripper was introduced that is driven by SMA actuators. The use of SMA offers many advantages as mentioned in the introduction and it results in a simple construction. However, the use of classical control theories leads to difficult problem as the feature of SMA is strongly non-linear, further each actuator has different features. In order to eliminate this problem we proposed a solution based on fuzzy logic algorithm. Our aim was not to construct a precision positioning system, but a robot gripper that can move in different trajectory using SMA wires. After the learning phrase we applied SVD reduction method that compressed the rule bases to 73%. Controlling one finger the SVD method has no important role, but in our future work when whole robot hand will be developed the reduction of the control rule base is inevitable. In conclusion the Fuzzy rule base could learn the proper feature of the whole system. Our future work is to develop a robot hand with five fingers.

ACKNOWLEDGEMENTS

The authors wish to thank the National Technical Research Found (OTKA: F 026127) for the financial support.

REFERENCES

1. C. Liang, C.A. Rogers: Design of Shape Memory Alloy Actuators, Journal of Mechanical Design, Vol. 114, June 1992, p. 223-230.
2. L.C. Brinson, M.S. Huang: Simplification and comparison of shape memory alloy constitutive models, Journal of Intelligent Material Systems and Structures, 1995
3. Duering, T.W. at al.: Engineering Aspects of Shape Memory Alloys, Butterworth-Heinemann Ltd, London 1990, ISBN 0-750-61009-3
4. Roger G. Gilbertson: Muscle wires project book, Mondo-Tronics, Inc. 1994, ISBN 1-879896-13-3.
5. Steven G. Shu et al.: Modelling of a flexible beam actuated by shape memory alloy wires, Journal of Smart Material and Structure, Vol. 6, Nr. 3, 1997, pp. 265-277
6. J. Rudas, J.F. Bitó, J.K. Tar: An Advanced Robot Control Scheme Using ANN and Fuzzy Theory Based Solutions. Robotika, 1996, Vol. 13, pp. 189-198. Cmbridge University Press.
7. K. Erbatur, O. Kaynak, J. Rudas: Fuzzy Identifier Based Inverse Dynamics Control for 3-dof Articulated Manipulator. Proceedings of the 23rd International Conference on Industrial Electronics, Control, and Instrumentation (IECON'97), Nov. 1997, New Orleans, USA, pp. 1052-1056.
8. J.K. Tar, J. Rudas, J.F. Bitó: New Prospects in the Adaptive Control of Robots Under Unmodeled Environmental Interactions. Proceedings of the 23rd International Conference on Industrial Electronics, Control, and Instrumentation (IECON'97), Nov. 1997, New Orleans, USA, pp. 1355-1360.
9. Yeung Yam, "Fuzzy approximation via grid points sampling and singular value decomposition", *IEEE Trans. System Man and Cybernetics to appear.*
10. P.Baranyi and Yeung Yam, "Singular value-based approximation with non-singleton fuzzy rule base" *7th Int. Conf. Fuzzy Systems Association World Congress* (IFSA'97) 1997 Prague pp 127-132.
11. P.Baranyi and Yeung Yam "Singular Value-Based Approximation with Sugeno Type Fuzzy Rule Base" *6th IEEE Int. Conf. on Fuzzy Systems* (FUZZ-IEEE'97) 1997 Barcelona, Spain
12. Yeung Yam, P.Baranyi and Chi Tin Yang "Reduction of Fuzzy Rule Base Via Singular Value Decomposition" *IEEE Trans. on Fuzzy Systems.* (submitted)
13. P.Baranyi, Yeung Yam, L.T.Kóczy "Multi Variables Singular Value Based Rule Interpolation" *IEEE Int. Conf. System Man and Cybernetics* 1997 (IEEE SMC'97) Orlando, Florida, USA, Vol. 3, pp. 1598-1360
14. Mihálcz István: New Shape Memory Alloy Power Circuits - RAAD'97 Cassino, Italy, 1997 Jun. 26-28, 459-462 oldalak. ISBN 88-87054-00-2

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Simulation of an Automatic Vehicle Transmission as a Mechatronic System

A. Haj-Fraj and F. Pfeiffer

Lehrstuhl B für Mechanik, Technische Universität München D-85747 Garching, Germany

This paper deals with the modelling of vehicle automatic transmissions. To get realistic predictions about the system behavior during the gear shift operations, the models of all drive train components are presented. Particular attention is paid to the detailed description of the hydraulically electronically controlled wet-clutches.

1. INTRODUCTION

Vehicle automatic transmissions (figure 1) represent highly complex technical systems. A very important issue of the development of automatic vehicle transmissions is an optimal control of the shift operations. A shift operation is denoted as optimal when the driver senses this event as comfortable.

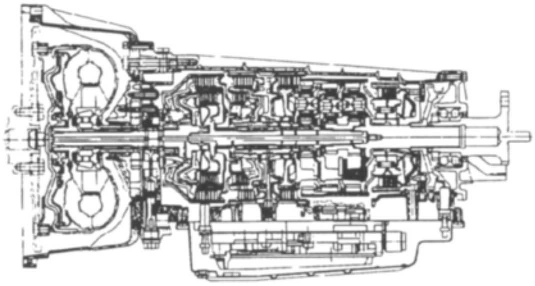


Figure 1: Five-Speed Automatic Transmission

The gear shift operations are carried out by controlling oil-immersed clutches and brakes. Normally during a gear shift operation two clutches are operating simultaneously, one clutch engaging and the other disengaging.

In [1] and [4] the fundamentals of powershifting in automatic transmission are presented by a rigid body model which includes quite simplified clutch and drive train models. The models used in these papers allow rough investigations on mechanical phenomena during powershifts, but is not convenient to get realistic simulations of the dynamic behavior of the drive train. A comprising model incorporating the drive train, the gear box, the vehicle, and the electronic control is developed for simulation and optimization of gear shift operations.

2. MECHANICAL MODEL

The decomposition of a technical system into subsystems allows a modular structure of the mechanical model and makes the numerical implementation in a simulation program very efficient. A drive train with automatic transmission consists in general of five main components: engine, torque converter, gear box, output train, and vehicle. Each component of the drive train can be considered as a rigid multibody system. The partitioning into single bodies is often given through their technical function. In case of elastic shaft, a discretisation of the body can be made by the stiffness and mass distribution. Thereby only torsional degrees of freedom are considered. The rigid bodies are coupled among each other and with the inertial environment by ideal rigid joints, clutches, and force elements.

2.1. Engine Model

For investigations on gear shift comfort the high frequency vibrations of the engine are not of great interest. The engine can therefore be modelled as a rotating rigid body. The drive train excitation caused by the engine is described by its torque which can be interpolated from a measured two-dimensional characteristic map as a function of throttle opening and the engine angular velocity.

2.2. Torque Converter Model

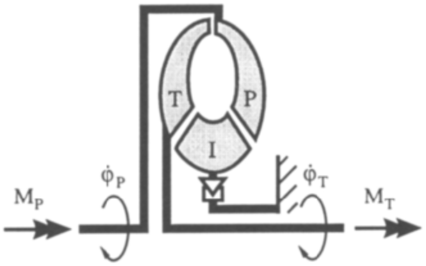


Figure 2: Hydrodynamic Torque Converter

The hydrodynamic torque converter consists of a pump (P) which is connected to the input shaft, a turbine (T) which is connected to the output shaft, and an impeller (I) which is pivoted on the housing through a one-way clutch (figure 2). Investigations on the vibrational behavior of the converter can be performed by a detailed model with four degrees of freedom. Since we are interested here just in the transmission behavior of this component we describe it as a force law. With the definition of the velocity ratio ν and and torque ratio μ (figure 3)

$$\nu = \frac{\dot{\varphi}_T}{\dot{\varphi}_P}, \; \mu(\nu) = \frac{M_T}{M_P} \tag{1}$$

and for given input and output angular velocity $\dot{\varphi}_P$ and $\dot{\varphi}_T$ the turbine torque can be calculated as

$$M_T = \mu(\nu) M_P \tag{2}$$

The pump torque itself can be calculated using a torque characteristic $M_{PC}(\nu)$ measured by a constant pump angular velocity $\dot{\varphi}_{PC}$

$$M_P = M_{PC}(\nu) \left(\frac{\dot{\varphi}_P}{\dot{\varphi}_{PC}} \right)^2 \tag{3}$$

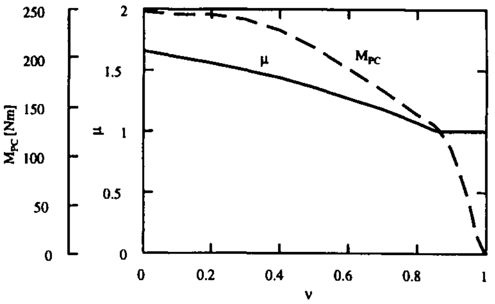


Figure 3: Hydrodynamic Torque Converter Characteristics

2.3. Gear Box Model

The gear box (figure 4) consists of three planetary sets which are connected to each other and to the gear housing by shafts, clutches, or one-way clutches.

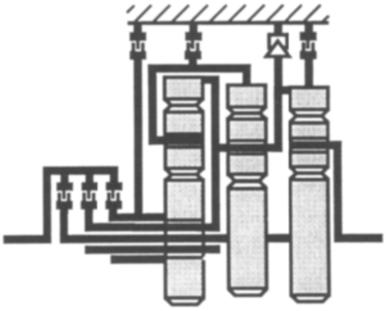


Figure 4: Gear Box Model

Planetary Set Model

Assuming a rigid mesh of tooth the kinematics of the planetary gear (figure 5) is described by the following conditions of pure rolling

$$\begin{aligned} \dot{\varphi}_S \cdot r_S &= \dot{\varphi}_P \cdot h \,, \\ \dot{\varphi}_R \cdot r_R &= \dot{\varphi}_P \cdot (h - 2 \cdot r_P) \,, \\ \dot{\varphi}_C \cdot r_C &= \dot{\varphi}_P \cdot (h - r_P) \,. \end{aligned} \tag{4}$$

where, $\dot{\varphi}_S$, $\dot{\varphi}_R$, $\dot{\varphi}_C$ are respectively the angular velocities of sun gear, ring gear, and carrier, and r_S, r_R, r_C are the corresponding radii. Eliminating the planet wheel velocity $\dot{\varphi}_P$ and the distance h from eqs. (4) we obtain

$$\dot{\varphi}_S \cdot r_S + \dot{\varphi}_R \cdot r_R - 2\dot{\varphi}_C \cdot r_C = 0 \tag{5}$$

Therefore the planetary gear can be described by two degrees of freedom.

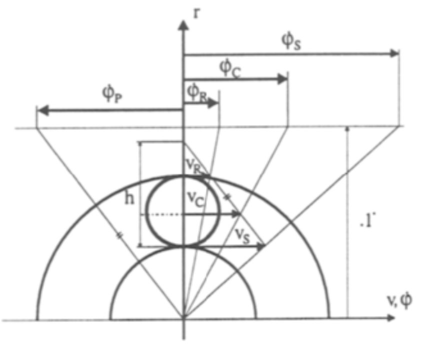


Figure 5: Planetary Gear Kinematics

Wet Clutches Model

A detailed model (figure 6) is used to describe the torque transfer in the clutch [3].

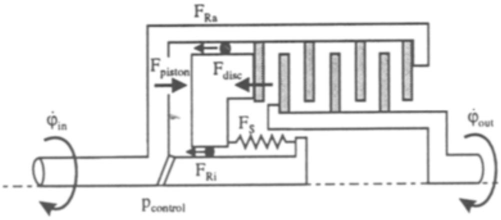


Figure 6: Multi Disc Wet Clutch

At the beginning of a clutch engagement there is an oil film between each two neighbouring discs. During the engagement the oil has to be squeezed out before friction surfaces have reached contact and the clutch can stick. We distinguish three phases during a clutch engagement: viscous slipping, mixed slipping and sticking.

The transition from viscous to dry friction is coupled with a smooth rise of the contact area in each pair of friction surfaces. In order to describe this transient phase we take the geometry and elasticity of the steel disc asperities into account.

The axial dynamical behavior of the piston can be described by its equation of motion

$$m_{piston}\ddot{x}_{piston} = F_{piston} - F_R - F_S - F_{disc} \ . \quad (6)$$

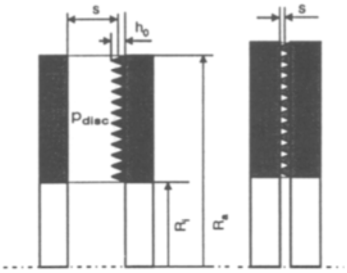


Figure 7: Friction Surface Model

where F_{piston} represents the force resulting from the pressure in the chamber behind the piston, F_R the friction force between piston and housing, F_S the return spring force, and F_{disc} the force between two neighbouring discs. By integrating eq. (6) with respect to time we obtain the piston position x_{piston} which gives the film thickness, s. Depending on s and the asperity's mean height h_0 we can then determine whether viscous or mixed friction occurs:

- $s > \frac{h_0}{2}$ viscous slipping: here the disc force F_{disc} results from the pressure in the oil film which can be obtained by solving the corresponding Reynolds equation of the squeeze process. The clutch torque is equal to the viscous torque that follows from the integration of the newton shear stress approach with respect to the friction surface A_R.
- $s < \frac{h_0}{2}$ mixed slipping: in this phase we have to deal with the asperity's elastic deformations [7]. After determining the contact surface and force in each pair of friction surfaces we calculate the dry torque. The oil pressure and viscous torque in the areas between the asperity's contacts can be calculated as described below. We obtain the whole disc force by summing the contact and oil forces and the whole clutch torque by summing the viscous and dry torque.

The sticking state occurs when the relative angular velocity in the clutch disappears. In this case the clutch torque represents a constraint torque.

One-Way Clutch

One-way clutches allow a relative motion in one direction and block it up in the other direction. Their use makes it possible to perform gear shift operations by steady comfort quality and low technical effort. The state of a one-way clutch can be described by its relative velocity $\dot{\varphi}_{OWC}$ which is defined as the difference of the velocities of the coupled components. Setting the free direction as the positive one the state of the one-way clutch is determined as follows:

- $\dot{\varphi}_{OWC} = 0 \Rightarrow$ one-way clutch blocks
- $\dot{\varphi}_{OWC} > 0 \Rightarrow$ one-way clutch unblocks

The case $\dot{\varphi}_{OWC} < 0$ can not occur because of the impenetrability condition.

2.4. Output Train Model

The output train consists of the cardan shaft, the differential gear, the output shafts, and the wheels.

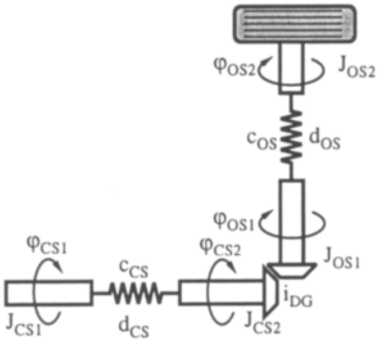


Figure 8: Output Train Model

Since we are interested in low frequency oscillations of the drive train we take into account the elasticities of the drive shaft and of the output shaft. The elasticity of the differential gear can be neglected here because of its very high stiffness. Consequently this submodel has three degrees of freedom (figure 8) $(\varphi_{CS_1}, \varphi_{CS_2}, \varphi_{OS_2})$. φ_{OS_1} follows from the kinematic constraint

$$\varphi_{OS_1} = \frac{1}{i_{DG}} \varphi_{CS_2} \tag{7}$$

2.5. Vehicle Model

The vehicle model (figure 9) used here takes into account the vehicle mass, the rolling friction, the driving resistance, and the tire elasticity and damping.

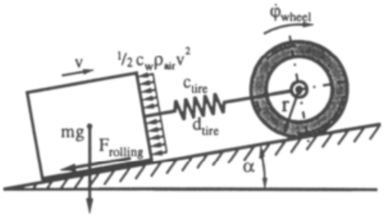


Figure 9: Vehicle Model

3. MATHEMATICAL MODEL

In order to investigate the dynamical behavior of the transmission system and its interaction with the other components during the gear shifting we consider the transmission system in the whole power train. The whole mechanical model is shown in figure 10.

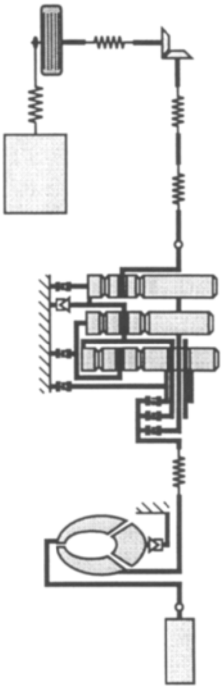


Figure 10: Drive Train Model

For the derivation of the drive train's equations of motion we consider the model without the gear shift elements. After the equations of linear and angular momentum of each submodel are formulated and transformed into the space of the generalized coordinates by the corresponding Jacobians the equation of motion of the drive train can be written as follows

$$\begin{aligned} \boldsymbol{M}\ddot{\boldsymbol{q}} - \boldsymbol{h} - & \sum_{i \in I_{wc}} (\boldsymbol{w}_{disc} F_{disc} + \boldsymbol{w}_{wc} M_{wc})_i - \\ & \sum_{j \in I_{owc}} (\boldsymbol{w}_{owc} M_{owc})_j = \boldsymbol{0} \ . \end{aligned} \tag{8}$$

The sets I_{wc} and I_{owc} consist of the n_{wc} numbers of the active wet clutches and the n_{owc} numbers of the active one-way clutches.

The vector of coordinates $\boldsymbol{q}$ represents the generalized but not necessary minimal coordinates of the system. For example if a clutch switches from slipping to sticking the number of degrees of freedom become smaller and the vector of minimal coordinates is no longer $\boldsymbol{q}$. The Matrix $\boldsymbol{M}$ represents the positive definite system mass matrix according to the vector $\boldsymbol{q}$. $\boldsymbol{h}$ is the vector of all active generalized forces without the contact forces in the gear shift elements. The contact forces F_{disc_i} in the friction surfaces, contact torques M_{wc_i} in the active clutches, and M_{owc_i} in the active one-way clutches are transformed into the space of of coordinates $\boldsymbol{q}$ by the constraint vectors

$$\begin{aligned} \boldsymbol{w}_{disk_i} &= \left(\frac{\partial h_i}{\partial \boldsymbol{q}}\right)^T , \\ \boldsymbol{w}_{owc_i} &= \left(\frac{\partial \dot{\varphi}_{owc_i}}{\partial \dot{\boldsymbol{q}}}\right)^T , \\ \boldsymbol{w}_{wc_i} &= \left(\frac{\partial \dot{\varphi}_{wc_i}}{\partial \dot{\boldsymbol{q}}}\right)^T . \end{aligned} \tag{9}$$

Due to the transitions in the clutches and the one way-clutches and the corresponding changes of the number of degrees of freedom, the equations of motion (8) have time variant structure [2]. Using a complementary formulation of the contact laws the system of equations of motion (8) can be transformed into a linear complementarity problem (LCP) [6].

The equations of motion and the LCP can be solved using numerical methods. Because of the time variant structure of the equations of motion, a determination of the switching points is necessary during the numerical integration.

4. GEAR SHIFT CONTROL

The gear shift in automatic transmission is realized by controlling wet clutches. The necessary control pressure is generated by a hydraulic device. The corresponding pressure controllers and magnet valves provide the clutch with the oil pressure. The control pressure depends on the kind of the gear shift, the turbine angular velocity and torque, the gear output angular velocity, oil temperature, etc. and is calculated in the electronic control device by linear interpolation from different pressure characteristic maps. Figure 11 shows a schematic curve of the pressure during the gear shift process [5].

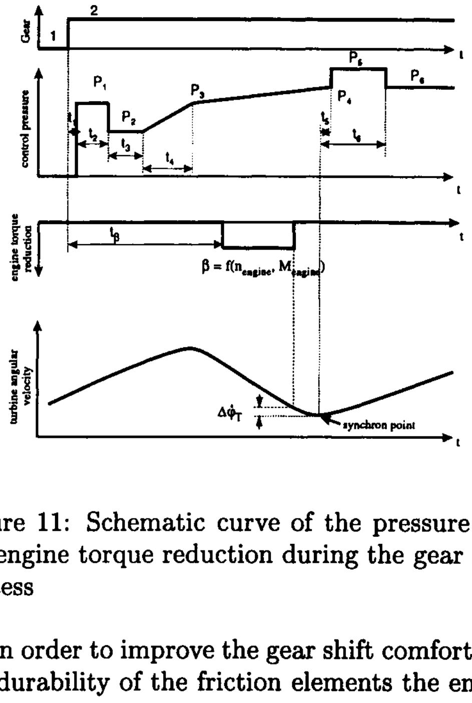


Figure 11: Schematic curve of the pressure and the engine torque reduction during the gear shift process

In order to improve the gear shift comfort and the durability of the friction elements the engine torque is reduced during the gear shift operation. The reduction factor β is determined in the electronic control device and transferred to the digital engine electronics by the controller area network. The torque is interpolated from corresponding characteristic maps as a function of the engine velocity and torque. A schematic curve of the engine torque reduction during the gear shift operation is shown in figure (11).

5. RESULTS

In this chapter we present some simulation and experimental results to verify the model of the drive train. As an example we consider a gear shift from the first to the second gear which is performed by engaging a wet clutch. During this operation a one-way clutch unblocks automatically depending on the torque development in the engaged clutch.

Figure 12 shows the torque and relative angular velocity in the one-way clutch and controlled wet clutch. After determining the switch point from the gear shift characteristic map, the clutch torque increases along with the control pressure. Simultaneously the torque in the one-way clutch decreases continously. At the moment when the clutch transfers enough torque to reverse the turbine angular velocity, the one-way clutch unblocks and its torque disappears. The disappearance of the relative angular velocity in the clutch implies the sticking of the clutch and means that the transmission ratio of the second gear has been established.

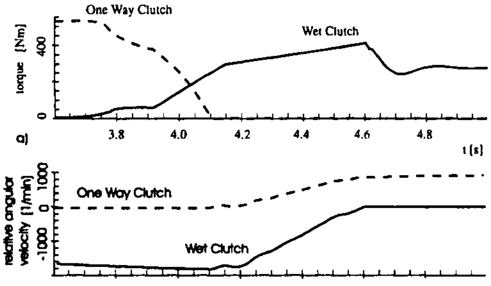


Figure 12: Torque and relative angular velocity in the one-way and wet clutch

To verify the presented model figure 13 shows a comparison of simulated and measured plots of the angular velocity of the turbine and the gear output shaft. The accordance of measurement and simulation confirms the mechanical model which represents a solid foundation for continuing research to develop a computer based optimal control of vehicle automatic transmissions.

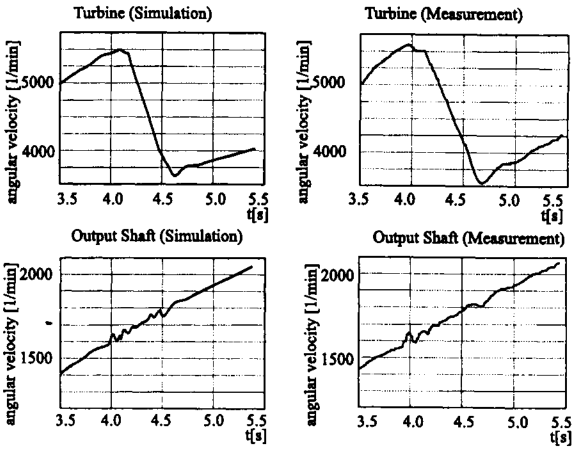


Figure 13: Comparision of simulation and measurement

References

[1] FÖRSTER, H.-J.: *Getriebeschaltung ohne Zugkraftunterbrechung.* Automobil Industrie, pp 60-76, 1962.

[2] HAJ-FRAJ, A.: *Dynamik der Schaltvorgänge in Automatikgetrieben.* VDI-Berichte No 1285, pp 359-373, 1996.

[3] HAJ-FRAJ, A.: *Modellierung der Schaltelemente in Pkw-Automatikgetrieben.* VDI-Berichte No 1323, pp 415-427, 1997.

[4] KRAFT, K.-F.: *Zugkraftschaltungen in automatischen Fahrzeuggetrieben.* Dissertation at the Universität Karlsruhe, 1972.

[5] KÜÇÜKAY, F.; RENOTH, F.: *Intelligente Steuerung von Automatikgetrieben durch den Einsatz der Elektronik.* ATZ Automobiltechnische Zeitschrift 96, No 4, pp 228-235, 1994.

[6] PFEIFFER, F.; GLOCKER, CH.: *Multibody Dynamics with Unilateral Contacts.* John Wiley Inc., New York, 1996.

[7] YAMADA, K.; TAKEDA, N.; KAGAMI, J., NAOI, T.: *Mechanisms of elastic contact and friction between rough surfaces, Wear 48, pp 15-34, 1978.*

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Resonance and quasi-resonance drives for start-stop regime

T. Akinfiev[a] and M. Armada[b]

[a] Mechanical Engineering Research Institute of Russian Academy of Sciences, M.Kharitonievskiy 4, Moscow 101830, Russia*

[b] Instituto de Automatica Industrial CSIC, Carretera de Camporreal, Km. 0,200. La Poveda 28500 Arganda del Rey, Madrid, Spain

Methods are discussed of increasing of effectiveness of machines working in start-stop regime. Results are given of analytical calculations, simulation and experimental research.

1. INTRODUCTION

Usually the increase of speed of machines can be achieved by an increase of capacity of a drive. However, there are many specific machines and gears which working elements perform a movement with a stop and stoppage (MSS), for example, various robots, cyclic automatic devices, rotary tables, classifying heads, sampling conveyors, automatic warehouses, automatic doors, scanning systems etc. All these machines are characterised by a constant alternation of acceleration and braking. The movement occurs in a non-constant-speed condition or makes up a significant part of a total time of movement. For such machines it is possible to increase substantially their effectiveness, that is, a speed of MSS machines can be increased several times with simultaneous decreasing of energy expenses.

Why just this type of machines has such reserves for effectiveness' increasing? The matter is that usually electromotors work effectively only with a speed close to a data-sheet value. In the systems under consideration, however, working elements (and therefore, drives) are changing their speed over a very wide range during the whole working process practically. That causes a low drive's effectiveness. In the paper two different possibilities of increasing of MSS machines effectiveness are considered.

2. CONSTRUCTION PRINCIPLES

2.1. Resonance drive

One of these two possibilities uses resonance ideas for creation of working movements of MSS machines' executive elements. It is well known from the theory of fluctuations that acceleration and braking in resonance oscillatory systems are performed at the expense of passive spring elements, where electromotor serves only for making up for friction losses in the system. The idea of using resonance shed new light on the problem of increasing of effectiveness of MSS machines. This idea had been developed first in [1], but the employment of electromagnet as energy resource did not allow to create real resonance MSS machines. Development of the idea had been connected further with the use of electromotors that allowed applying an effective control [2].

The resonance effect allows, in comparison with traditional drives, the increase of quickness of a machine, and, at the same time, the reduction of the power input. This is done without a complication of the construction of the machine on the whole. The special technique of designing has allowed creating various MSS systems with resonance properties, so named Resonance Drives (RD) [3].

---

* Present address: Instituto de Automatica Industrial CSIC, Carretera de Camporreal, Km. 0,200. La Poveda 28500 Arganda del Rey, Madrid, Spain. Fellowship CSIC.

The simplest type of RD is intended for revolving or progressive movements of a mobile link 1 (see fig.1) with a stop and fixation in extreme positions. The resilient member 2 between the carriage and the support structure is adjusted so that the mobile link should be in a state of balance when it is at the centre of the distance between its extreme positions. When the mobile link is at one or another extreme, retaining device 3 automatically retains it. In its initial position, the mobile link is at one or the other extreme. To start the motion of the mobile link, the movable part of the retaining device is disengaged from the stationary part. With the mobile link released, the resilient member moves it towards its target position by accelerating it over the first half of its path and decelerating it over the second half of its path. When the mobile link arrives in the target position it is automatically retained again by the retaining device. A drive motor is connected to the mobile link by means of a working wheel 4 and serves only to compensate for friction losses.

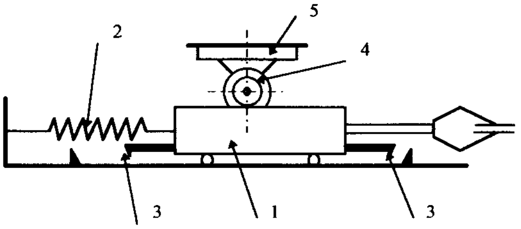


Figure 1. Resonance drive.

Some other kinematic circuit designs have been elaborated [3], in particular, drives with non-linear resilient members including springs with «negative stiffness», lock-less drives and so on. These drives allow making movements with stops not only in extreme positions, but also in any intermediate point. They also allow making a transfer from every point of positioning to any other, and so on. However, in any case the main principle is unaltered - during a movement inertia forces are compensated by passive resilient members and motor serves only to make up for energy losses in the system. That allows using a low-power motor and, in compare to traditional drives, lowering energy expenses more than 10 times and increasing quickness simultaneously.

2.2 Quasi-resonance drive

Another possibility is connected with the use of special non-linear reduction gears with smoothly modifiable gear ratio between electromotor and working element. Besides, it is desirable a gear ratio to be tended to infinity when close to extreme position. In this case the intensive acceleration - braking regime of working element would be maintained, and working element would be automatically held in extreme positions due to self-braking effect in reduction gear. When approaching a middle point a gear ratio should decrease. It is desirable to arrange this decrease, as so the speed of a rotor would be practically constant at significant changes of speed of working element.

It will be shown further that dynamic characteristics of the drives with non-linear reduction gears are very similar to those of RD. Therefore, the drives with non-linear reduction gears have been called quasi-resonance drives (QD).

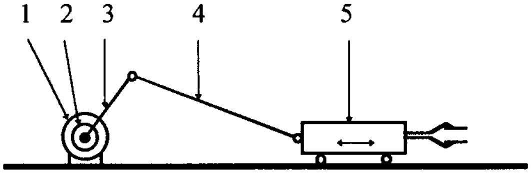


Figure 2. Quasi-resonance drive.

On fig. 2 the simplest kinematics circuit of QD drive is represented. The electromotor 1 is connected with reduction gear 2, which has constant gear ratio k_1. On an output shaft of this reduction gear a crank 3 is fixed, which is hinged through the rod 4 with working element 5. It is assumed that working element can be moved only in a horizontal direction; it stops in extreme positions. If a length of the crank 3 is taken as l, the value of movement of working element is $2l$. Horizontal positions of the crank correspond to stop positions of working element. In these positions the crank automatically locks working element, so as it cannot move having an effect of external forces on it. The moving of working element from any positioning position to another one can be executed by rotation of the crank both in positive and negative direction.

3. ANALYTICAL CALCULATIONS

Systems of differential equations, which describe dynamic properties of RD&QD, have a considerable number of non-linearities (in some cases up to 20). In this connection a special method for analytical calculation of dynamic properties of the drives has been elaborated. This method is based on the ideas of harmonic linearisation [4] and had been used already for solving problems of dynamic analysis and synthesis of different RD [5,6]. The method assumes both getting the first approximation and refining of solutions obtained. The method of getting of the first approximation in dynamic analysis and synthesis problems of QD is presented below. Analytical refining of solutions obtained has not been made because numerical methods of integration of full equations has been used for getting solutions refined

Setting up equations of movement, let mass of working element m be substantially more than mass of the other mobile parts of the machine. Supposing that time of stoppage does not practically affect the dynamic properties of a system, instead of a movement with a stop and stoppage in a real system let us investigate a special simulating system, which performs a movement with stops and without stoppages. We write then equation of movement of working element in the form:

$$m\ddot{x} + b\dot{x} + N\,\mathrm{Sign}\,\dot{x} = Mk_1k\,, \tag{1}$$

where x is a coordinate of working element, measured from middle position; b and N - coefficient of viscous and force of dry friction, respectively; M - moment of DC motor; k - gear ratio in non-linear reduction gear. Assuming that a length of a rod is much more than a length of a crank, we can write a relation $k = \dfrac{1}{l\,\mathrm{Sin}\,\varphi}$, which connects this gear ratio with an angle of a crank's turn. Supposing that electromagnetic process in electromotor run considerably faster than electromechanical, we have:

$$M = i\gamma\,, \tag{2}$$

where - γ is a constant parameter of the DC motor.

The magnitude of a current in electromotor's armature can be calculated by the formula:

$$i = \frac{U - E}{R}\,, \tag{3}$$

where U is a value of a voltage that feeds electromotor and depends on the value of a crank's turn angle, R is an armature's resistance. E is electromotive force of a self-induction in (3), calculated by the formula:

$$E = \dot{\alpha}\gamma\,, \tag{4}$$

where $\dot{\alpha}$ is angular speed of the armature. A relation connects this speed with angular speed of a crank $\dot{\varphi}$:

$$\dot{\alpha} = k_1\dot{\varphi} \tag{5}$$

For complete description of the considered simulating system, a type of the law of controlling has to be established. Taking into consideration periodic and symmetric character of movement, it is possible to choose elementary law of electromotor's controlling in the form:

$$U = \begin{cases} U_1 & 0 \prec \varphi \prec 0.5\pi \\ U_2 & 0.5\pi \prec \varphi \prec \pi \\ U_1 & \pi \prec \varphi \prec 1.5\pi \\ U_2 & 1.5\pi \prec \varphi \prec 2\pi \end{cases} \tag{6}$$

Bearing in mind plenty of non-linearities in the considered system of differential equations, as well as periodic and even oscillatory character of movement of working element in simulating system, let us employ method of harmonic balance for getting solutions of these equations.

From the kinematics it is obvious that amplitude of fluctuations of working element is equal to crank's radius. Therefore we seek a solution in the form:

$$x = -l\,\mathrm{Cos}(t\omega)\,, \tag{7}$$

where $\omega = Const$ is a circular frequency of fluctuations of working element.

According to the method of harmonic balance, we replace non-linear functions by harmonically linearized approximations. The system has a property of a filter, so let us consider only first harmonic component. Taking into account orthogonality of trigonometric system of functions we find final relations, describing dynamic properties of considered simulating system:

$$
\begin{aligned}
ml\omega^2 &= \frac{4k_1\gamma}{\pi Rl}\left(U_1 - U_2\right) \\
bl\omega + \frac{4N}{\pi} &= \frac{k_1\gamma}{lR}\left(U_1 + U_2\right) - \frac{\gamma^2 k_1\omega}{lR}
\end{aligned}
\qquad (8)
$$

System of finite equations (8) is well known in oscillation theory. Usually, first of these equations is called «skeleton curve»; it defines own properties of oscillation system. Second equation is usually called «balance energy curve». This curve has the property that in its every point the energy that feeds the system is equal to the energy that is dissipated by the system in the process of a motion. The point of intersection of these two curves corresponds to desired solution.

It has to be noticed that curves (8) are completely identical to corresponding curves that had been obtained as a result of analysis of RD [5,6]. Actually, the difference of voltages U_1-U_2 multiplied by constant coefficient in QD corresponds to stiffness of spring element in RD, and the sum of voltages U_1+U_2 in QD corresponds to double voltage of motor's power supply in RD.

The availability of final equations permits to solve problems of dynamic analysis and synthesis for the system under consideration. In the considered system the amplitude of fluctuations of working element is determined by geometrical size of the crank and is known beforehand. It enables to use as a second unknown parameter any other coefficient, for example, U_2.

Let us put a problem of seeking of such a reduction gear $k_1 = k_1{}^*$, that moving of a working element on a given distance during a given time requires a minimal magnitude of voltage U_1. From a system of equations (8) we find:

$$k_1{}^* = \frac{l\sqrt{\pi mR\omega}}{2\gamma} \qquad (9)$$

When having such a reduction gear, voltage U_1 should satisfy to a following condition:

$$\frac{U_1^2}{R} = \frac{9}{16}\pi ml\omega^3 \qquad (10)$$

Thus, the only parameter (10), which needs to be taken into account when selecting electromotor, is a power of thermal losses in armature of braked motor. This result - a coordination of losses in a source of energy and in useful load - correlates well with results, received as a part of a study of RD for start-stop regime [7].

4. CONTROL SYSTEM

It is important, that RD&QD work effectively only when their parameters satisfy specific tuning relations. However, during a real work process some of parameters of a drive can change. Thus, when using these drives in robotics, a mass of a moving part can change. It can be also substantial that during some movements, the gravity helps moving, and during others it impedes moving. Friction forces can also change, some casual disturbing effects can occur, and so on. All that can be not only responsible for less effective work of drives but sometimes can cause a total loss of drive's normal operation. Therefore, for a maintenance of a reliable work of a RD&QD in real conditions, it is necessary to have a special adaptive control system, which could supply automatic tuning of a drive if different constructive parameters change.

When elaboration of such a system one has to take into consideration a substantial fact that a system has to work in a real time regime. Usually, while using RD&QD the whole process of moving takes a relatively little time (0.2 - 0.6 s). Thus, it is clear that a control algorithm could not be based on the utilisation of a prolonged calculation procedure. One more factor influences a selection of a control algorithm: the fact that a lot of different kinematic circuit designs of the drives for different purposes have been elaborated by now. It would be a good idea to elaborate such an algorithm that a constructed on its base control system could control all RD&QD with the same efficiency without preliminary re-tuning, i.e. could be universal.

On the basic of use of principles, discussed in [2] a special algorithm of adaptive control based on ideas of self-training and symmetrization, has been elaborated. It satisfies all conditions enumerated.

Let us note that real values of speed depend on mass (or on moment of inertia) of a mobile link that can change from move to move and are not known at the point of beginning of movement. But ideal tuning of a system takes place when a mobile link comes to the extreme position with a zero speed. Such a movement is symmetric comparatively to a middle position of a mobile link. If we know a law of movement on the first part of a trajectory, then we would easily maintain symmetrical law of movement on its second part. For making that a movement's trajectory is divided into two parts: passive part - from initial position to middle point, and active part - from middle point to extreme position. In the simplest case a motion on passive part of trajectory is carrying out without any feedback, only feeding a constant voltage to a motor. At this time a mobile link begins to move due to the law, which practically corresponds to own properties of a system (with given parameters). During this movement a record of current coordinate values along with corresponding speed values is being made. After that, on active part of trajectory a symmetrization is carried out of a movement with drive's help. The idea is that if on passive part of a trajectory a motion begins from initial position with a zero speed, then on active - symmetric - part it will end with a zero speed in a positioning position, just as we wanted to get. For this purpose a feedback with respect to speed is introduced on active part of trajectory. In a feedback circuit a controller can be introduced. It minimises a deviation of actual speed value in every point of the second trajectory part from speed value in corresponding point (symmetric point relative to middle position) of the first trajectory part, i.e. it minimises a function $f = f(\dot{x}_x - \dot{x}_{-x})$.

From traditional control methods' point of view, such an approach reminds a training regime that is used widely in practice. However, a principle difference is that in our case training is made not before working process, but directly during working process. That allows using the same control system for different types of RD&QD.

A natural restriction of the control system has to be taken into consideration: a motor power has to be big enough to compensate disturbance energy. This restriction is especially important because the drive motor reacts against disturbances only on the second half of the trajectory. In the described control system information about some parameters of the system available before beginning of the movement, is not used. The use of such information increases significantly control system's possibilities (but makes it less universal).

5. SIMULATION

On the basis of packages MATLAB-SIMULINK special programs have been elaborated for simulation of work of RD&QD with universal system of adaptive control. Typical result of simulation is given on fig. 3, where the first series of the curves on phase plane corresponds to a drive without any control system (dotted and dash-dotted lines), and the second one - to a drive with a control system (solid curve).

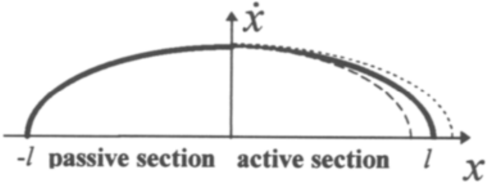


Figure 3. Control algorithm.

Results of simulation of RD&QD are shown on fig. 4 (coordinate as a function of time) and fig.5 (phase plane). It is evident that dynamic properties of RD (dotted lines) and QD (solid lines) are very similar though that is achieved differently. In RD passive spring elements are used, in QD special reduction gears are employed with changing gear ratio. These reduction gears provide electromotor's operation in the most effective regime when motor's speed is practically constant though speed of mobile link is changing over a wide range.

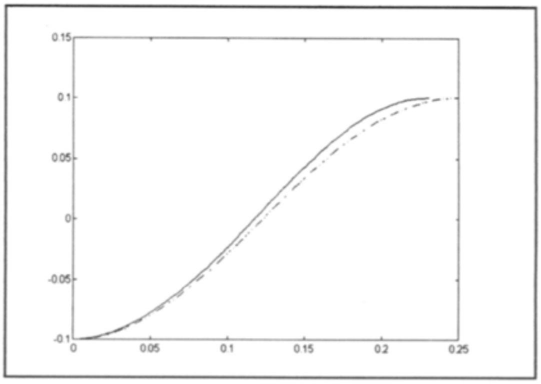


Figure 4. Law of motion

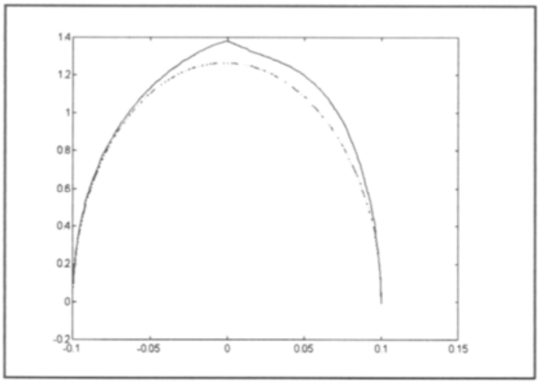


Figure 5. Phase plane

The results of simulation have shown that alteration of some constructive parameters of a drive such as friction force, step magnitude, spring stiffness as well as establishment of casual disturbing actions of reasonable intensity does not disrupt normal operation of a drive with adaptive control.

6. EXPERIMENTS

Experimental researches of resonance robot MARS [3] have shown that with the load 0.5 kg this robot provides the rise of arm on 50 mm, extending of arm on 200 mm and turn of arm on 90^0 in 0.2 s having drive power only 4 W.

Resonance portal robot with the load 10 kg executes horizontal displacement on 800 mm and vertical displacement on 300 mm in 0.5 s having drive power 90 W.

Special resonance block [3] for rotation of a robot's grip of 20 kg mass with a load 12 kg provides grip's rotation with step 90^0, and speed 180 ^{0}/s having drive power 60 W.

Traditional facilities for executing the above described operations have 10 times more power motors and also have much less quickness.

For experimental comparison of traditional drives with RD&QD, special models of RD&QD have been made for horizontal displacement of a leg of walking robot. These drives have the same motor of 15 W power as base traditional [8] drive. They provide displacement of 5-kg load on 100 mm. However, under these conditions RD&QD have 10 times higher speed than a base drive.

Experiments have shown that the use of universal adaptive control system provides reliable work of the drives and that the control system works in plug-and-play regime.

7. CONCLUSIONS

It has been shown that RD&QD have similar dynamic characteristics and provide increase of speed of machines working in start-stop regime decreasing simultaneously energy expenses more than 10 times. Elaborated universal adaptive control maintains reliable drives' operation without operator intervention and without preliminary work in training regime.

ACKNOWLEDGEMENTS

The authors would like to thank EC Commission for financial support of this work in the frame of the INCO-COPERNICUS «ESSIDAC» Project.

REFERENCES

1. Ridderstrem G. Mechanical Hand, USSR Patent No 568346 (1975).
2. Akinfiev T. Method of Controlling Resonance Hand, US Patent No. 4958113 (1990).
3. Akinfiev T. Resonance Drives with Adaptive Control. Proc. ASCE Engineering Mechanics Conference, Florida, USA, 947 (1996).
4. Besekerskii V., Popov E. Theory of Automatic Control Systems, Moscow, 1975.
5. Akinfiev T. et. al. About the Influence of a Mass of the Load on the Resonance Manipulator Dynamics. Ac. Sci. USSR, Mashinovedenie, 1, (1985) 37.
6. Akinfiev T., Pozharinskii A. Dynamics of Manipulation System with the Resonance Drive. Ac. Sci. USSR, Mashinovedenie, 6 (1984) 10.
7. Akinfiev T. Resonance Manipulation Systems with Electric Drives. Ac.Sci.USSR, Mashinovedeniye, 6 (1983) 18.
8. Armada M. et. al. Four-legged Walking Test-bed, First IFAC International Workshop on Intelligent Autonomous Vehicles, Hampshire, UK, 8 (1993).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Hardware Oriented Modeling and System Identification: Breaking the Barrier between Design and Analysis

Mike Donnelly, Analogy, Inc., Beaverton, OR and Joachim Langenwalter, Analogy Europe

Hardware oriented modeling and simulation, combined with system identification techniques, improves communication between traditional electrical and mechanical designers and control systems analysts. Specialists in power electronics, electromagnetic actuators, mechanisms, etc., can model and simulate their designs at the physical implementation level. They can then capture the analytical equivalent model, and check it against the expectations of the controls analysts, sometimes with surprising (and informative!) results.

The design environment for complex motion control systems often includes a barrier between the system analysis function and the hardware design function. At the system level, controls analysts deal with abstract characteristics, using linear analysis (e.g. state-space, root locus, etc.) or "block-diagram" based simulation methods. System design considerations include:

- Stability and margins
- Performance issues (e.g. speed of response, settling time, tracking error, etc.)
- Operating modes and fail-safe strategies

At the hardware level, design engineers select parts, define circuit topologies, create mechanical drawings or otherwise define the physical components of the system. At this level, critical issues include:

- Part performance and rated capability
- Effects of tolerance or dimensional variation
- Manufacturing methods and overall cost of production

There are significant differences in these two "views" of the same design, in terms of objectives, methodologies and terminology. These differences create a communication barrier between the design functions.

Yet often performance improvement, problem solving and cost reduction is possible only when the barrier is removed and these functions are well coordinated.
For example, when several hardware design groups (e.g. electronics, power, actuators, mechanisms, hydraulics, etc.) are contributing to a system design, each group attempts to meet a subsystem specification that was defined at the system level. But interaction effects between these subsystems can sometimes get lost in the specification process. These interactions can affect the overall system dynamics and impact performance. Usually these problems are not discovered until integrated system test, long after many of the "easy-fix" windows are closed.

This article describes an effective process that identifies and helps solve these problems in the early stages of the design. A hypothetical design example of an antenna positioning system is used to illustrate the method. Using hardware oriented simulation, coupled with system identification techniques, hardware level design activities are tightly linked back to the system analysis. This "bottom-up" information flow is essential for successful integration of complex, multi-technology systems.

*Footnote to article:
All simulation models and results are shown in SaberSketch™ and SaberScope™, part of the SaberDesigner™ simulation environment from Analogy. All root locus plots and system identification results are shown in MATLAB® from the MathWorks. Data passing between MATLAB and Saber is provided by SaberLink™, also from Analogy.

Motion Control Example
Spacecraft Antenna Positioning System

Concept Development and Implementation

System designers often start with a basic concept, sketched out in rough form, which includes the major components of the system. Figure 1 is a simple sketch of the antenna positioning mechanism. The design topology and initial estimates of parameters are based on the system performance objectives and physical constraints, (e.g. the response and settling times required, the antenna feed mass, a single polarity bus voltage of +28vdc, etc.). This design concept is a motor driven rack-and-pinion mechanism with a return spring. The configuration is fail-safe; upon loss of electrical power, the spring will pull the antenna feed-horn back to its "home" position.

A simple hand analysis of the "plant" (motor and mechanism pair) dynamics is also shown in Figure 1. The differential equations are converted into transfer function form, so that major characteristics, such as natural frequency and damping, can be predicted.

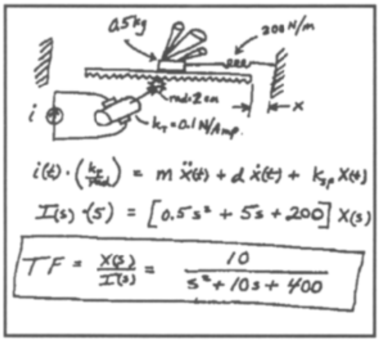

0.5kg
200 N/m
i
rad=2cm
x
$k_T = 0.1$ N/Amp.
$i(t) \cdot \left(\frac{k_T}{r_{rad}}\right) = m\ddot{x}(t) + d\dot{x}(t) + k_{sp}x(t)$
$I(s) \cdot (5) = [0.5s^2 + 5s + 200]\, X(s)$
$TF = \frac{X(s)}{I(s)} = \frac{10}{s^2 + 10s + 400}$

Fig 1.

This form is also needed for a block-diagram simulation model, as shown in Figure 2a.

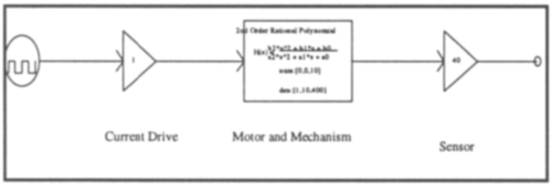


Fig 2a.

Note that in transforming the system description to the transfer function form, the relationship to the physical hardware becomes obscured. For example, the values of spring rate and antenna feed mass (200 N/m and 0.5 kg respectively) do not appear directly in the coefficients of the transfer function. This is the beginning of the barrier. Hardware designers will not be able to relate to this model of their design, at least not in terms of the parts and parameters they are concerned about. For example, the effect of variations in spring rate, due to tolerance, may be difficult to track through these transformations. The fact that there is a spring is not even obvious!

A more natural approach for the mechanism designers is shown in the hardware oriented simulation model in Figure 2b.

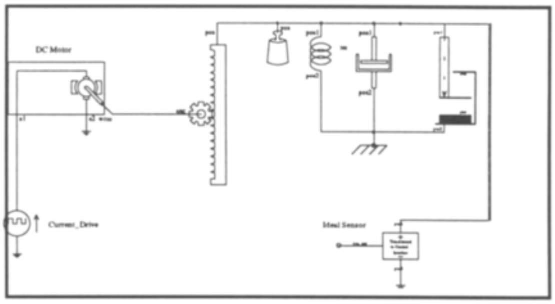


Fig 2b.

Note that all of the components, such as the motor and spring, are connected as they are in the physical system. Component parameters, such as torque constant and spring rate, are simply entered unchanged on the appropriate element. The simulator has the concept of conservation (e.g. the sum of the forces must equal zero). It can assemble the entire system of equations based on the individual parts' force vs. motion characteristics, and the way the parts are connected or assembled. Additional effects, such as friction of the rack, backlash in the pinion, etc., can be added simply by connecting the appropriate element on the schematic.

The simulated responses to a step change in current are shown in Figure 2c.

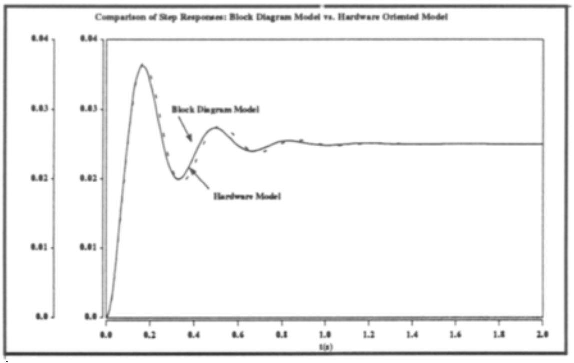


Fig 2c.

Note the slight difference between these responses. This difference is due to the fact that, in addition to the mass of the antenna feed-horn, the motor and drive shaft inertia also contribute to the total inertia of the system. This was overlooked in the system level model, but shows up in the hardware oriented model, because it is a specific parameter of the motor element and is a well known factor to the mechanism designers. Though not very significant in this case, this finding illustrates the potential advantage of maintaining a tight coupling between design levels.

Closed-loop System Design

The open-loop system step response is very lightly damped, with high overshoot and long settling time. Overshoot is not acceptable for this position control design, because a large step transition at the end of travel would crash the feed into the mechanical stop. Also, this open-loop configuration would be subject to position errors caused by parameter variations in the plant. For example, when the spring gets cold and stiff during eclipse, the feed position vs. current relationship would change.

The control system designers can work to overcome these concerns. Closed-loop feedback control can be used to minimize the under-damped response and improve robustness to parameter changes. The root locus for uncompensated proportional feedback is shown in Figure 3a.

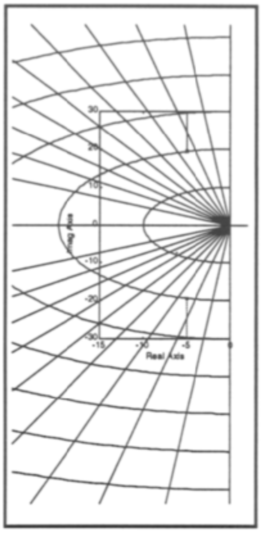


Fig 3a.

Note the system poles move straight out, parallel to the imaginary axis, implying the system would get more oscillatory. Lines of constant damping fan out from the origin; closer to the vertical axis is less damped. Lead/lag compensation is then applied. The new root locus (Figure 3b) shows improved damping while maintaining response speed, even for a relatively high DC gain of 100.

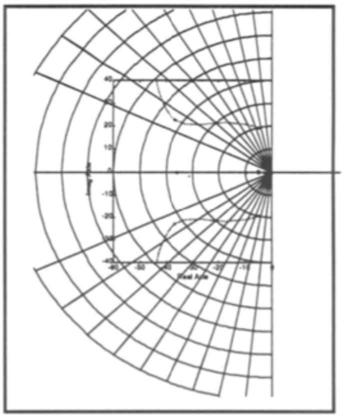


Fig 3b.

The controller design is verified with a block-diagram simulation, as shown in Figure 4.

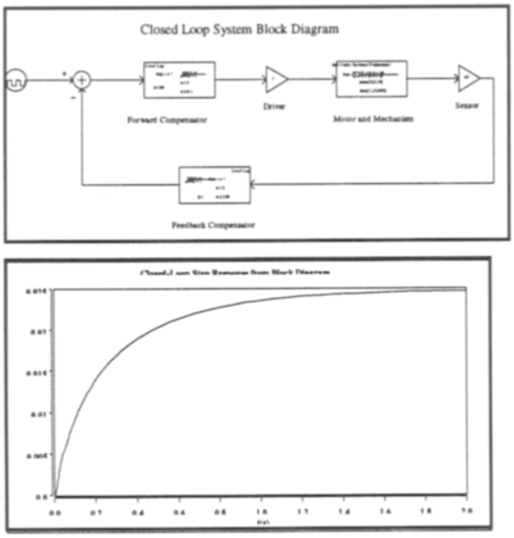


Fig 4.

The new step response shows good speed with no overshoot, and a very small steady-state error because of the high loop gain. This high loop gain also provides robustness to parameter variations.

Detailed Design Implementation... A Problem is Uncovered!

The hardware oriented simulation model of the entire system is shown in Figure 5, including the electronic implementation of the compensators as well as the switching motor driver.

Two features are important to note. First, because it is a switch mode design, the motor driver is not linear, and therefore cannot be analyzed directly using automated linearization methods.

In addition, the system designers specified a motor drive with current control, 1 Amp/Volt gain. But this particular motor driver design is actually voltage control, with a 1 Amp/Volt current gain at steady-state only. The actual motor dynamics, which includes back-EMF generation, contributes an additional pole to the system transfer function, and this leads to some unexpected results.

These are observed in the step response, also shown in Figure 5. The system is unstable, with significant oscillation during the transition. This was not predicted by control system analysis, which was based on the idealized system model.

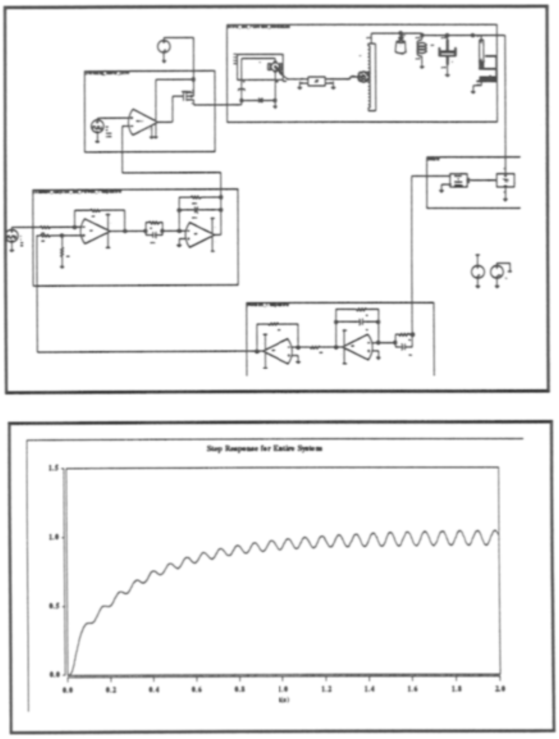


Fig 5.

System Identification Pinpoints the Problem

The designers may not immediately recognize their assumptions about the plant dynamics are incorrect, but they may suspect an interaction between the motor drive and the motor, because that section of the design is non-linear and not easy to analyze. However, using the hardware oriented model, a simulated characterization test is set up to identify the dynamics of the combined drive circuit, motor and mechanism portion of the system, as shown in Figure 6.

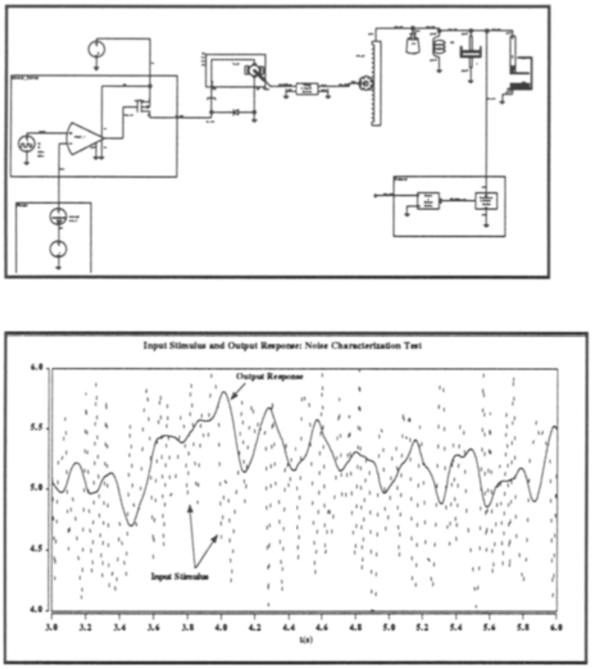

Input Stimulus and Output Response: Noise Characterization Test
Output Response
Input Stimulus
6.0
5.5
5.0
4.5
4.0
3.0
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.8
5.0
5.2
5.4
5.6
5.8
6.0
t(s)

Fig 6.

MATLAB is used to generate a noise stimulus signal, which is passed to Saber. The resulting input/output response is passed back to MATLAB, where the System Identification Toolbox is used to estimate a transfer function model for this portion of the system.

Note in Figure 7b, that the fit is quite good for a 3rd order polynomial model. The 2nd order model (Figure 7a) shows a very poor fit to the noise response. This result is unexpected, based on the originally assumed dynamics.

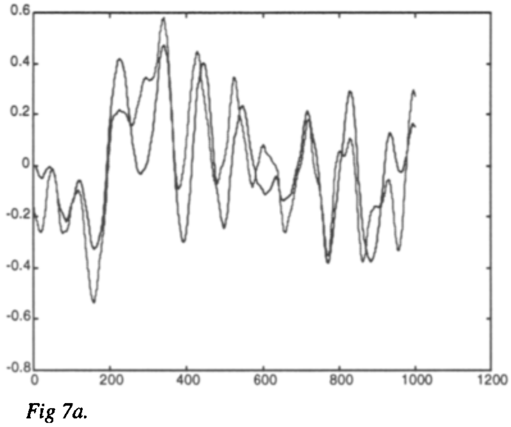


Fig 7a.

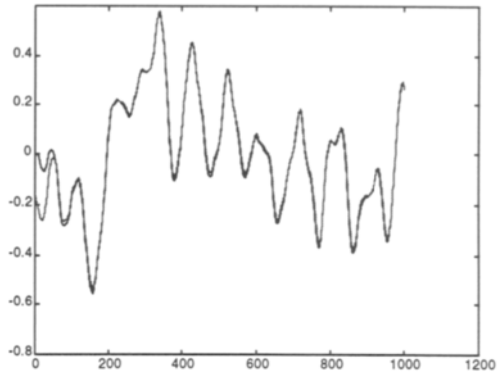


Fig 7b.

The difference is due to the extra pole from the voltage-mode (instead of current-mode) operation of the motor driver. Thus, using this system identification technique together with the hardware modeling and simulation capability, the "smoking gun" was found!

Compensator Redesign

Once the actual plant dynamics are identified, a new root locus analysis is performed. In Figure 8a, the root locus has branches that move quickly into the right half-plane (i.e. the unstable region), and that at a gain of 100 the system is nearing instability.

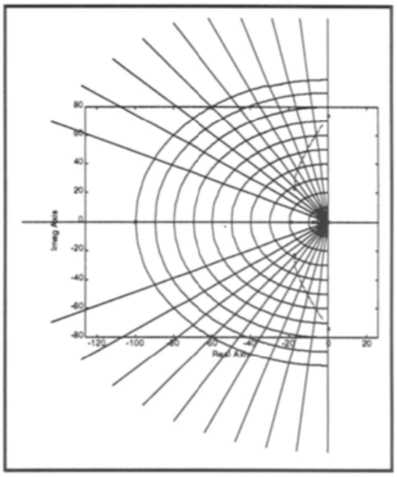


Fig 8a.

Next, the controller is redesigned, moving one of the compensator poles out farther into the left half-plane. The new root locus, shown in Figure 8b, avoids the right half-plane until much higher gains are reached.

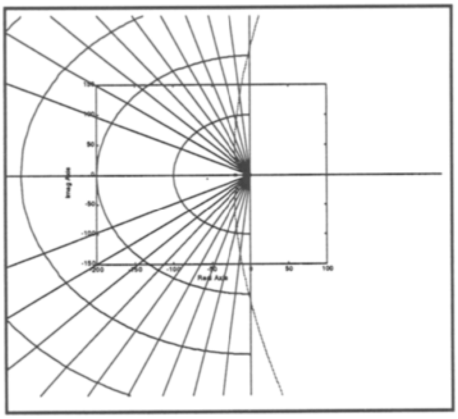


Fig 8b.

Verify the Solution and then Commit to Hardware

Implementing the change in the compensator pole frequency simply requires selecting an alternate capacitor value. In Figure 9, the capacitor is changed from 2uF to 200nF, and the full system step response simulation is re-run. The result is greatly improved performance, with no oscillation. The designers can feel confident that the system will now perform as expected, and they are ready to build hardware.

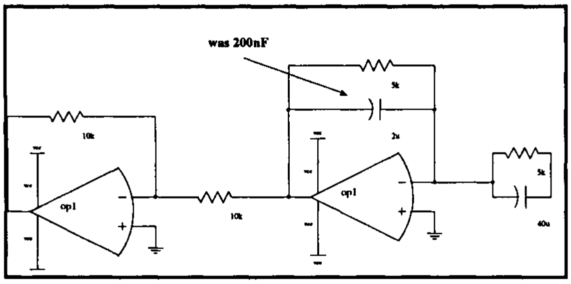


Fig 9a.

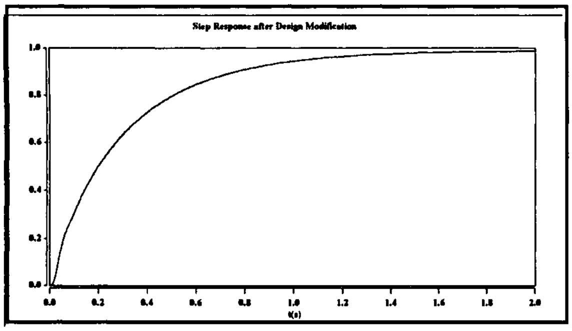


Fig 9b.

The alternative to this process is to find these interface problems at integrated hardware test. The system identification methods are equally applicable at that time, but schedule and cost impacts may be quite high. These impacts may be unavoidable, because analytical understanding is usually required to solve the problem. The approach described using simulation methods can identify the problem much earlier in the design cycle, and avoid an expensive delay in the project.

References

Lennart Ljung, "System Identification Toolbox Users Guide", The MathWorks, Inc.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A high-precision positioning servo-controller using non-sinusoidal two-phase type PLL

L. Wang[a] and T. Emura[a]

[a]Department of Mechatronics and Precision Engineering, Tohoku University,
Aoba-ku, Sendai, 980-8579, Japan

Most servomechanisms use encoders whose output signals are 90° phase-shifted sinusoidal waves for detecting angular position. These 2-phase signals are not true sinusoidal waves and their distortion makes the positioning accuracy worsen. Therefore, the authors proposed non-sinusoidal two-phase type PLL (Phase-Locked Loop) to increase positioning accuracy of servomechanisms. To obtain high-precision detection of angular position, an iterative learning method to compensate the wave distortion was tried, and a high-resolution and high-precision positioning controller that can use low-resolution encoder was obtained. This paper discusses the proposed method and gives experimental results.

1. INTRODUCTION

Positioning servo systems usually use encoders to detect linear or angular position. High-precision machines need high-resolution encoders. However, the maximum speed of ordinary high-resolution encoders is too low to detect high-speed motion due to the limitation of response frequency. Therefore, in order to detect the high-speed motion, a method that interpolates a low-resolution encoder is usually used. However, the conventional interpolators can not be used under the strong electrical noise, which is generated by high-power PWM (Pulse Width Modulation) servoamplifiers of large NC (Numerically Controlled) machines. To solve this difficult problem, T. Emura proposed a two-phase type PLL (Phase-Locked Loop) method[1], and the authors applied it a high-speed NC gear grinding machine. Because two-phase type PLL uses 90° phase-shifted sinusoidal waves, its dynamic characteristics are significantly improved compared with that of conventional PLL, and the excellent performance under strong noise situation was confirmed from many experiments[2,3].

Encoders of high-accuracy sinusoidal waves have complicated structure and can not rotate at high speed. On the other hand, the output waves of low-accuracy encoder are distorted and they are not true sinusoidal waves. Since the two-phase type PLL mentioned above requires high-accuracy sinusoidal waves, distorted sinusoidal waves induce large error to the interpolation. Therefore, the authors tried to use non-sinusoidal two-phase type PLL to interpolate the encoders that have these distorted waveforms. Because non-sinusoidal two-phase type PLL uses the same distorted waveforms as those of encoder, it is able to interpolate the encoder signal with high accuracy even though the encoder generates distorted waveforms. However it is necessary to know precisely the waveforms of encoders. Even if we simply use a more precise system, it is difficult to obtain high-resolution interpolator because the detection is sensitive to electrical noise. In this study, a successive approximation approach was used to detect waveforms of encoders. This method does not require precise system but can realize a much more precise interpolation even using encoders whose waveforms are distorted. A servomechanism using the interpolator of non-sinusoidal two-phase type PLL was implemented. Experiments of servo controller using proposed method were carried out and the effectiveness was verified. This paper discusses the non-sinusoidal two-phase type PLL and a positioning servo-controller constructed by using it. The computer simulation and experiments are also presented.

2. INTERPOLATION USING NON-SIN]USOIDAL TWO-PHASE TYPE PLL

2.1. Two-phase type PLL

Ordinary PLL[4], as shown in Figure 1(a), contains three basic components: a phase detector (PD), a loop filter (LF) and a voltage controlled oscillator (VCO).

The notations are as follows.

θ_i : phase of input signal
θ_o : phase of output signal
a : amplitude of input signal
b : amplitude of output signal of VCO
$f(s)$: transfer function of loop filter
K_v : gain of VCO
y_d : output of phase detector
e : output of loop filter

Input and output of the loop are sinusoidal signals $a\sin\theta_i$, $b\cos\theta_o$. Generally, the phase detector carries the following operation.

$$\begin{aligned} y_d &= a\sin\theta_i b\cos\theta_o \\ &= \frac{1}{2}ab\{\sin(\theta_i-\theta_o)+\sin(\theta_i+\theta_o)\}\,, \end{aligned} \tag{1}$$

where $\sin(\theta_i-\theta_o)$ is a low-frequency term and $\sin(\theta_i+\theta_o)$ is a high-frequency term [4]. The high-frequency term $\sin(\theta_i+\theta_o)$ is removed by a low-pass filter included in the LF. (1) is simplified as follows.

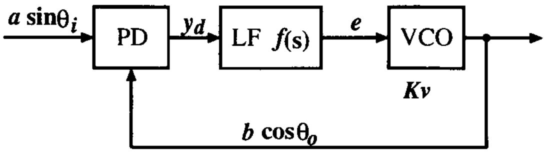
$$y_d = \frac{1}{2}ab\sin(\theta_i-\theta_o) = \frac{1}{2}ab(\theta_i-\theta_o) = k_d\phi\,, \tag{2}$$

where, $\phi = \theta_i - \theta_o$ is the phase difference and $k_d = ab/2$ denotes the gain of phase detector. When the phase difference ϕ is kept small by the feedback loop, the loop is in locked state. If the phase difference ϕ exceeds ±90°, the loop would cause cycle-slip or fall out of lock.

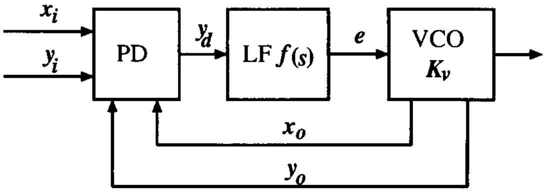
The cutoff frequency of loop filter must be set sufficiently low to remove the high-frequency term. However, the loop filter also determines the dynamic characteristics of the loop. When the interpolator needs wide bandwidth and high-response frequency, the cutoff frequency must be set high enough. Otherwise the interpolator could not track the input signal. In particular, if it is necessary to track an input signal that may change the sign of ω_i (the sign of ω_i represents the phase direction), ω_i and ω_o become zero when the shaft-speed of encoder passes the zero-speed state. This means that the ordinary PLL cannot be used, because the cutoff frequency of the loop filter cannot be set to zero.

Two-phase type PLL uses two-phase sinusoidal waves for input signals and signals generated from VCO as follows.

$$\left.\begin{aligned} x_i &= a\cos\theta_i \\ y_i &= a\sin\theta_i \end{aligned}\right\}, \qquad \left.\begin{aligned} x_o &= b\cos\theta_o \\ y_o &= b\sin\theta_o \end{aligned}\right\}. \tag{3}$$



(a) PLL.



(b) Two-phase type PLL

Figure 1. Block diagram of PLLs.

Phase detector for two-phase signals takes the following calculation.

$$\begin{aligned} y_d &= x_i y_o - y_i x_o = ab\sin(\theta_i-\theta_o) \\ &= ab\sin\phi \approx ab\phi. \end{aligned} \tag{4}$$

Comparing (4) with (1), we find that there exists no high-frequency component in the output of the two-phase type phase detector. This means that phase detector of two-phase type PLL detects phase difference independently of input frequency. Therefore, the performance on dynamic range can be improved.

2.2. Two-phase type PLL for non-sinusoidal signal

When the signals are not sinusoidal, the PLL can also lock up, when the function $y_d(\phi)$ is monotonous with respect to ϕ for all θ_i. Consequently, we can still use two-phase type PLL for interpolator if the following condition is satisfied.

$$\begin{cases} \{k\phi - h(\phi)\}h(\phi) > 0, & \phi \neq 0 \\ h(\phi) = 0, & \phi = 0 \end{cases} \tag{5}$$

This means that, if phase difference ϕ is detected it can be controlled to 0 by the feedback loop. If we use the same waveforms as those of encoder in VCO, when phase difference ϕ is controlled to 0, signals from encoder and VCO are coincident with each other. Thus the phase of VCO can be regarded as that of encoder and be detected from VCO.

For example, if the outputs of encoder are triangular waves, the characteristics of phase detector can be ob-

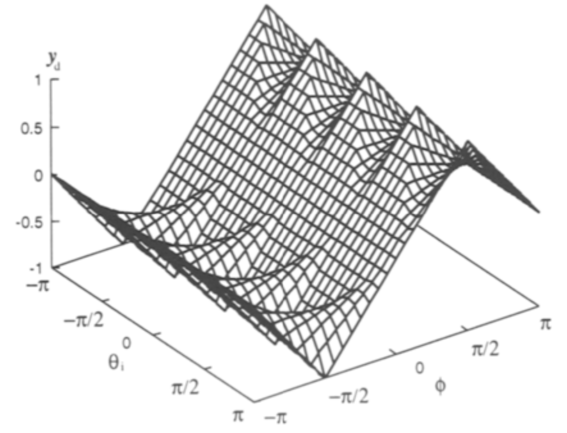


Figure 2. Output of phase detector with triangular waves.

tained as shown in Figure 2. From Figure 2 we know y_d is proportional to ϕ when ϕ is small. Therefore, the low-cost non-sinusoidal wave encoders can also be interpolated with high precision by two-phase type PLL. The two-phase type PLL with non-sinusoidal waves is called non-sinusoidal two-phase type PLL in this paper.

2.3. Interpolator using non-sinusoidal two-phase type PLL

Figure 3 shows a block diagram of an interpolator using non-sinusoidal two-phase type PLL. VCO for the two-phase type PLL consists of a V/F(Voltage to Frequency) converter, a reversible binary counter, two ROM's, and two D/A converters. The V/F converter controlled by the output signal of loop filter generates pulse train with the frequency proportional to the input voltage. The output pulses of V/F converter are counted with the reversible counter, and the output value of counter is fed to the address input of ROM's. In the two ROM's, tables of waveforms are written. These waveforms are the same as those of encoder. Thus the frequency of the two-phase signal is proportional to the input voltage of V/F. When the phase difference ϕ is controlled to zero, the output pulses of V/F converter can be used as interpolated pulses of encoder. As the results, the interpolated pulses are generated from V/F converter or a computer reads the reversible counter of VCO directly.

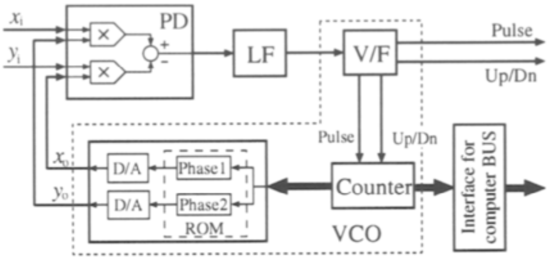


Figure 3. Block diagram of interpolator using two-phase type PLL.

3. SERVO CONTROLLER USING NON- SINUSOIDAL TWO-PHASE TYPE PLL

The authors tried to use the interpolator of non-sinusoidal two-phase type PLL to a high-precision servomechanism. As mentioned above, it is required to use the same waveforms in VCO as those of encoder. Otherwise, the difference between the waveforms results in large interpolation error. In fact, PLL is able to lock up with different waveforms. However, ripples appear at the output of phase detector even y_d is small. This means that we cannot control y_d to 0 and obtain phase angle of input signal precisely. The waveforms of encoder can be detected using a high-precision rotary table. But the system becomes complicated and it is impossible to make the accuracy higher than the rotary table. Therefore, the authors proposed a method that makes waveforms of VCO approximate encoder's waveforms successively by correcting waveforms of VCO iteratively.

3.1. Detection of waveforms using successive approximation method

Figure 4 illustrates the block diagram of proposed method, where vectors $\boldsymbol{r}$ and vectors $\boldsymbol{r}$ represent input signal and signal of VCO, that is, $\boldsymbol{r}_i = (x_i, y_i)$ and $\boldsymbol{r}_o = (x_o, y_o)$. The two-phase type PLL is constructed in computer by software for convenience of correction of waveform tables and other parameters. Encoder signals are input through A/D converters. In Figure 4, $\Delta\boldsymbol{r}$ denotes the waveform difference between VCO and encoder, that is $\Delta\boldsymbol{r} = \boldsymbol{r}_i - \boldsymbol{r}_o$. If the input is periodic signal, the generated signal from VCO is periodic as well when PLL is in lock state. Thus the waveform difference vector $\Delta\boldsymbol{r}$ is periodic with the same period. The correction is carried out based on

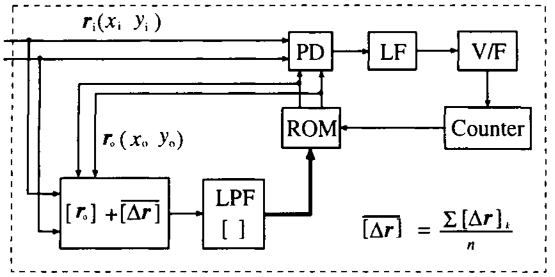

$r_i(x_i\ y_i)$
PD
LF
V/F
ROM
Counter
$r_o(x_o\ y_o)$
$[r_o]+[\overline{\Delta r}]$
LPF
[ ]
$[\overline{\Delta r}] = \frac{\Sigma[\Delta r]_k}{n}$

Figure 4. Block diagram of software two-phase type PLL.

the period. After PLL is locked, r_o and Δr are stored in the memories of computer. At the end of one cycle of input signal the correction $[r_o] + [\Delta r] \rightarrow [r_o]$ is carried out, where [ ] indicates the table of one period of data. When PLL locks up again the error appearing at the output of phase detector is reduced with the replaced waveform table $[r_o]$. By repeating the correction, the difference between waveforms is reduced and the waveforms of VCO approximate to those of encoder successively.

Figure 5 gives the procedure of successive approximation of waveforms, where initial waveforms in ROM tables are supposed to be sin and cos. The shaft of encoder is controlled at a constant velocity. If using rotary encoder, it is only to mount the encoder to a speed controlled motor. After PLL locked up, the error Δr is memorized for a number of period. These data are averaged as follows.

$$[\overline{\Delta r}] = \sum_{k=1}^{n} \frac{[\Delta r]_k}{n}, \tag{6}$$

where n indicates the number of period in which the data are stored. The average is carried out among corresponding data in each period. By averaging these data a table $[\overline{\Delta r}]$ of one period is obtained. And then waveforms of VCO are corrected using $[\overline{\Delta r}]$ as described above. This correction is carried out repetitively until the error reaches a preset accuracy or dose not decrease.

The average is an important operation in the procedure. This is because error which is not synchronous with those of encoder can be removed by the average if the number n is big enough. In practice, the output of encoder is always affected by noise, shaft vibration and rotary fluctuation. Such disturbances are rarely synchronous to encoder's signal. The influence of disturbances can be suppressed, this means that the proposed method dose not require complicated hardware setup such as precision rotary table. For example, as the error associated with rotary fluctuation has not influence in the final results, the rotary speed of encoder shaft is not required to be controlled accurately.

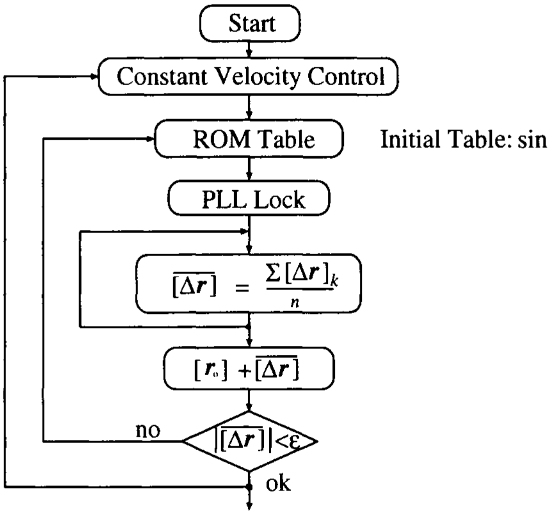
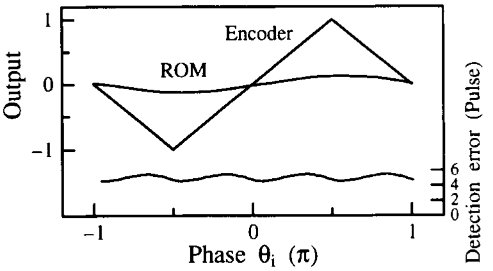

Start
Constant Velocity Control
ROM Table
Initial Table: sin
PLL Lock
$[\overline{\Delta r}] = \frac{\Sigma[\Delta r]_k}{n}$
$[r_o]+[\overline{\Delta r}]$
no
$|[\overline{\Delta r}]|<\varepsilon$
ok

Figure 5. Procedure of successive approximation of waveforms.

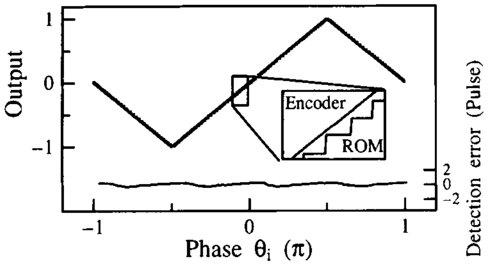
To avoid some accidental phenomena that disturbance is synchronous with encoder's signal, we can control encoder's shaft to rotate at different speed. Because the frequency of encoder output is changed, disturbance synchronized before speed changed will not synchronize at a new speed.

3.2. Simulation Using Triangular Waves

Simulation was carried out to verify the proposed method. Figure 6 illustrates some of simulated results, where only one phase of encoder's signal and one phase of VCO's are shown. In the simulation a two-phase triangular wave is used for encoder's output and sinusoidal one for initial waveform of VCO. The interpolating rate is 128. PLL is locked with these initial waveforms, but error is large as shown in Figure 6(a), where pulse indicates the interpolated pulse. After 20 times of iterative correction the error is reduced as shown in Figure 6(b). The error can be reduced to a very small value by increase the times of iteration. The behave of error during iterative correction is shown in Figure 6(c).



(a) Initial waveforms.



(b) Waveforms of 20th approximation.

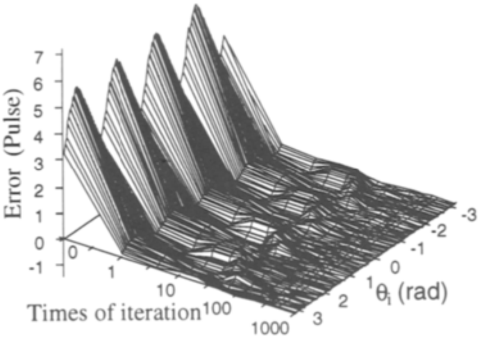
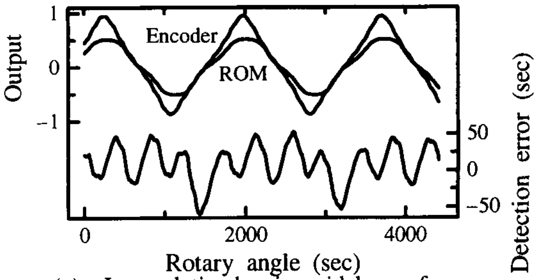


Figure 6. Simulated results.

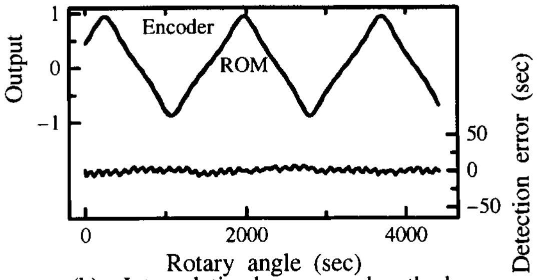
From the results of simulation, it is clear that (1) PLL can lock up with different waveforms and (2) the error due to waveform difference becomes very small after sufficient times of iteration. The second one shows the convergence of the iterative correction.

4. EXPERIMENTS

A rotary encoder is used for interpolation. The encoder has an output of 750 slits per revolution, that is its resolution is 1728 sec (arc second)[1]. The experiment is to interpolate it by a rate of 128 to increase the resolution to 13.5 sec. The shaft of encoder is driven with a servomotor that is speed controlled by a PI controller. For verification of interpolation accuracy of proposed method, a high-resolution encoder is mounted to the same axis. The output pulses from the high-resolution encoder are accumulated in a counter and fed to a computer. By comparing the output of the counter and interpolator, we can investigate the accuracy of interpolator.



(a) Interpolation by sinusoidal waveforms.



(b) Interpolation by proposed method.

Figure 7. Experimental results of detecting waves of encoder.

4.1. Detecting waves of encoder

Figure 7 is experimental results. The waveforms of encoder are distorted from sinusoidal waveforms. When we use sinusoidal waves to interpolate them, the detection error was large. This situation is shown in Figure 7(a) where the error is 112 sec pp(pick to pick). However the detection error was reduced to 13.5 sec pp by using proposed method.

4.2. Servo control using non-sinusoidal two-phase type PLL

The authors tried to use the interpolator of non-sinusoidal two-phase type PLL to a high-precision positioning servomechanism that uses traction drive. Figure 8 shows the block diagram. An encoder of

---

[1] sec is used to stand for second of rotary angle and s for second of time in this paper.

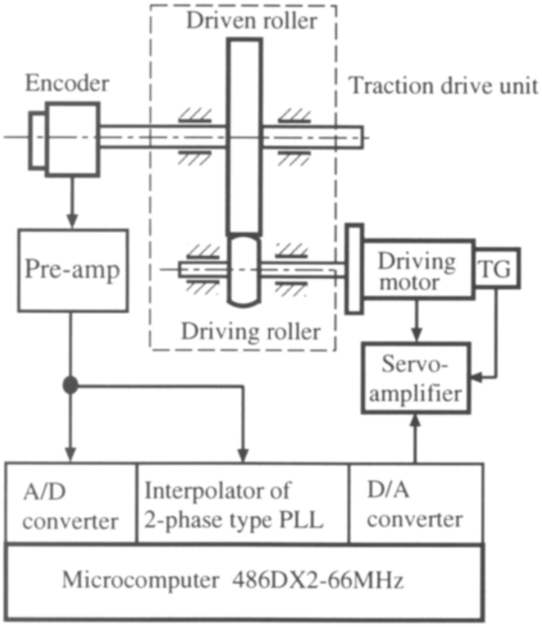


Figure 8. Implementation for experiments.

5 000 slits per revolution is used for detect the angular position of output axis. The original resolution is 259.2 sec. The interpolation rate of interpolator is 256 sec and the detection resolution of position is about 1 sec. Figure 9 shows the experimental results of step response. The amplitude of angular step input is 130 sec. From the results, we known the output axis tracks the input well and the ringing of tracking error in static state is 30 sec pp which is a small value compared with the original resolution of the encoder.

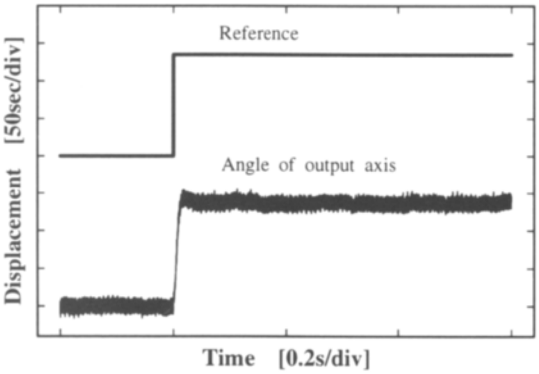


Figure 9. Step response of positioning servomechanism.

5. CONCLUSIONS

In this study, the authors tried to use low-cost and low-resolution encoders for high-precision positioning servomechanisms. Because output signals of low-cost encoders are distorted, non-sinusoidal two-phase type PLL was proposed for interpolating phase angle of encoder output with high accuracy, and a successive approximation method was proposed for increasing interpolation accuracy. A servo-controller that uses this non-sinusoidal two-phase type PLL was designed and fabricated. Experiments were carried out to confirm the effectiveness of this new type PLL. From the experiments, the following conclusions were concluded.

(1) Non-sinusoidal two-phase type PLL is useful for high-precision interpolation of distorted output waveforms obtained from low-cost encoders.
(2) To accurately detect the phase angle of distorted waveforms, a successive approximation method was proposed, and its effectiveness was confirmed by the experiments.

6. ACKNOWLEDGEMENTS

A part of this research works was supported by the Grant-in-Aid for Scientific Research of Japan Ministry of Education and the Foundation for Promotion of Advanced Automation Technology. We would like to express thanks to these grants.

REFERENCES

1. T.Emura, “A Study of A Servomechanism for NC Machines Using 90 Degrees Phase Difference Method,” Prog. Rep. of JSPE, pp. 419-421, 1982.
2. T. Emura, A. Arakawa and M. Hashitani, “A Study of High Precision Servo-Spindle for Hard Gear Finishing Machines,” The International Conference on Advanced Mechatronics, pp.427-432, 1989.
3. T. Emura, L. Wang and A. Arakawa, “A Study on a High-Speed NC Gear Grinding Machine Using Screw-Shaped CBN Wheel, ” Jour. of Mechanical Design, Trans. of ASME, Vol. 116, No. 12, 1994, 1163-1168.
4. F. M. Gardner: Phaselock Techniques. New York: Wiley, 1966.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Modeling And Adaptive Control Of Permanent Magnet Synchronous Motors Using Multilayer Neural Networks

Ayhan Albostan[a] and Muammer Gökbulut[b]
[a]Erciyes University Department of Electronic Engineering Kayseri / TURKEY

[b]Fırat University Faculty of Technical Education, Department of Electronic and
Computer Education Elazığ / TURKEY

Abstract

In this study, identification and adaptive tracking control of a permanent magnet synchronous motor (PMSM)--widely used in high-performance drives because of high torque, power density, and power factor--is implemented using Multilayer Neural Networks (MNN). A problem generally encountered in industrial drives is the identification of nonlinear, unknown motor-load dynamics. The purpose of this study is to identify nonlinear motor-load dynamics using MNN, and to control the motor with neural networks so that it can follow an arbitrarily selected reference speed or position. Hence, MNNs are used for the emulation and control of PMSM. The proposed algorithm is simulated using the MATLAB program, and the simulation results show that MNN is more effective than conventional adaptive control methods in nonlinear system modeling and control.

1. INTRODUCTION

Variable-speed drives constitute an important part of motion control in industrial production. Improvements in magnetic materials, semiconductor power devices, and control methods have resulted in the widespread use of permanent magnet synchronous motors in drive systems. In comparison with asynchronous and synchronous motors of known type, permanent magnet synchronous motors offer many advantages, such as high torque, power density and power factor. Magnetizing current drain from the stator has been prevented by using a permanent magnet in the rotor, and compared to asynchronous motors, the power factor and efficiency have been improved. Rotor losses have been reduced using a permanent magnet instead of a brush-ring system, DC power source and rotor winding as in the case of synchronous motors, and important advantages, such as high torque, high power and ease of maintenance, have been obtained over asynchronous and synchronous motors of the same size (rating). Further, by the use of various control methods, self-controlled PMSMs are able to display the speed and torque characteristics of direct current (dc) motors, which is a desirable advantage in drives. In addition, due to improvements in permanent magnet structures and PMSM design, these motors are preferred in systems with high-performance drive requirements, such as robotics and automated machine tools [1]. Such drive systems are required in industrial automation applications where sensitive motions are called for, and tracking accuracy should not change even when the load conditions, inertia, and parameters of the system do so. Hence, the control method should be adaptive, accurate, and easily applicable [2].

PID and optimal control methods, which require the derivation of an accurate mathematical model for the system to be controlled, cannot be used effectively in high-performance systems. Adaptive control methods, such as variable-structure, self-tuning, and model reference methods, do not require that a mathematical model of the system's dynamics be known. The dynamic system model is determined on the basis of input-output information of the system subject to control, and is generally a linear model, but may be refreshed every few sampling periods. Modeling and controlling a nonlinear system by linearizing it under variable working conditions does not yield good performance, and it is impossible to definitely determine the stability of a system whose dynamics cannot be modeled. In recent years, various adaptive control methods have been applied to

PMSMs, and motor performance has been improved by developing certain algorithms. While these methods are effective, they have been developed using approximations such as the constancy of linearization parameters, and generally call for intensive computations. Furthermore, the hardware realization of adaptive control methods is difficult [3,4,5].

Neural networks, which aim to artificially model the principle of functioning of the human brain, have begun to be used extensively in the field of control in recent years for the solution of the problems described above, in addition to their use in many other fields such as image and pattern recognition and signal processing. In particular, because of the ability of MNN to approximate any nonlinear function with the desired accuracy, various problems have been investigated in the field of control and nonlinear system modelling by the use of MNN, and the effectiveness of adaptive control via neural networks has generally been demonstrated [6,7,8]. In connection with high-performance drive systems, inverse models of dc and brushless dc motors have been obtained using MNN, and the use of neurocontrollers has been realized [9,10]. In the present study, PMSMs have been modeled in real time using MNN with static and dynamic learning algorithms, and their indirect adaptive control has been realized.

2. MULTILAYER NEURAL NETWORKS (MNN)

Although different definitions are possible, an MNN is a parallel distribution processor structurally capable of storing and subsequently using empirical information [11]. This definition renders an MNN analogous to the human brain in two respects:

- The MNN gains information after a learning process.
- Weightings between the neurons are used to store information.

As in the human brain, the "neurons" or processing elements are not units powerful enough individually to calculate or represent information, yet the interconnection of these elements in various ways produces the MNN and provides a powerful computing ability. A processing element can be treated, in short, as a unit that produces an output by passing the weighted sum of inputs through a nonlinear function . The nonlinear activation function of the processing elements is a linear function, a sign function, or a bidirectional (-1,1) or unidirectional (0,1) sigmoid function.

Neural networks are generally classified as single-layer or multilayer and with feed forward or recurrent. In the control field, multilayer neural network structures with feed forward and external recurrent have attracted attention, and various applications have been investigated. MNNs are important in the control field and have attracted widespread attention for the following reasons [6,7]:

- Provided there are sufficient sigmoidal processing elements in the hidden layer, it has been demonstrated that a 3-layer neural network consisting of input, hidden, and output layers can approximate any nonlinear function with the desired accuracy. This provides a significant advantage in the control of nonlinear systems.
- The back-propagation algorithm based on gradient methods widely used in the control field is the basic learning algorithm of multilayer neural networks.
- Ability to tolerate an error is high due to parallel distributed structure.
- A neural network can be trained using measurements performed on a system, and the trained neural network has the ability to generalize on data not used during training.

Feedforward and recurrent multilayer neural networks—which have a greater ability to approximate a given function than other MNNs—have been used in this study. If the inputs of an MNN are independent of the outputs, i.e., the information flow in the neural network is from input to output, this is called a feed forward neural network structure. Neural network inputs can be influenced by outputs—a situation generally encountered in control systems—and this is known as a recurrent or dynamic neural network. Externally recurrent MNNs are shown in Figure 1.

In Figure 1, W^s_{ji} is the weighting on the j^{th} processing element of layer s from the i^{th} output of layer s-1, x is the nx1 input vector, v is the px1 hidden-layer output vector, y is the mx1 output vector, and x_0, v_0 are polarization inputs. The input-output relationships of a 3-layer neural network with linear output layers are given in Equation (1).

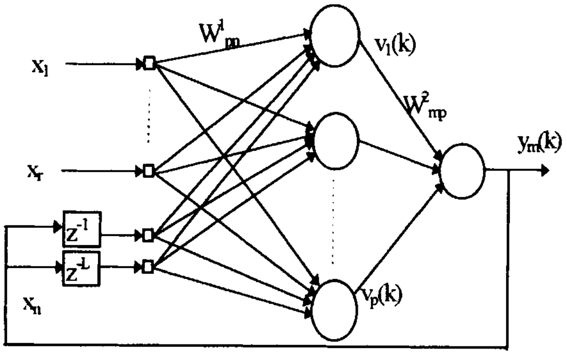


Figure 1. Externally Recurrent 3-Layer Neural Networks.

$$u_i = \sum_{l=0}^{n} W^1_{il}.x_l \,, \quad v_i = \varphi_i(u_i) \quad i=1,\dots p$$

$$y_j = \sum_{i=0}^{p} W^2_{ji}.v_i \quad j=1,\dots.m. \tag{1}$$

3. BACK-PROPAGATION LEARNING ALGORITHM

The most important role for MNNs in the field of control derives from their ability to accurately determine the input-output relationship of a system. Accordingly, the learning process of a neural network comprises the determination of the neural network coefficients by an optimum method with the aim of approximating the input-output function of a system. The principle of gradient methods—most widely used for the optimization of parameters in a system—is based on determining the partial derivatives of a chosen performance criterion with respect to the parameters. In neural networks, determination of the gradient of the performance criterion with respect to weightings is called the back-propagation algorithm. An accurate gradient of the performance criterion with respect to weights is the static back-propagation algorithm found by taking normal partial derivatives in feed forward neural networks, and the dynamic back-propagation algorithm found by taking ordered partial derivatives in recurrent neural networks [12]. The instantaneous learning algorithm which is the approximation of batch learning and is required in adaptive control because of its suitability to learning in real time. The sum of the squares of output errors is generally used as the performance criterion. Accordingly, with d_j as the desired output, the error and performance criterion in the k^{th} sample are:

$$e_j(k) = d_j(k) - y_j(k) \quad j = 1,2,\dots.m.$$

$$J(k) = \frac{1}{2}\sum_{j=1}^{m} e_j^2(k) \tag{2}$$

MNN weights are renewed with the partial derivative of the performance criterion with respect to a weight in the MNN.

$$\Delta W^s_{ji}(k) = \alpha \frac{\partial J(k)}{\partial W^s_{ji}(k)} \tag{3}$$

Consequently new weightings to be applied to the weights in the MNN are, with α representing the learning ratio:

$$W^s_{ji}(k) = W^s_{ji}(k-1) + \Delta W^s_{ji}(k) \tag{4}$$

For the MNN with linear output layer, the partial derivative of the performance criterion with respect to a weight in the hidden layer is:

$$\frac{\partial J(k)}{\partial W^1_{il}(k)} = \frac{\partial J(k).\partial e_j(k).\partial y_j(k).\partial v_i(k).\partial u_i}{\partial e_j(k).\partial y_j(k).\partial v_i(k).\partial u_i(k).\partial W^1_{il}}$$

$$= \varphi'(u_i)\sum_{j=1}^{m} -e_j(k).W^2_{ji}(k).\,x_l(k) \tag{5}$$

Because feedback exists in dynamical systems, it may be quite difficult and involved to calculate the partial derivatives of the performance criterion. Ordered partial derivatives have been taken as an important mathematical method in finding the partial derivatives of such systems, and the dynamic back-propagation algorithm has been found by applying them to neural networks. From the partial derivative of the performance criterion:

$$\frac{\partial J(k)}{\partial W_{ji}(k)} = \frac{\partial J(k).\partial y_j(k)}{\partial y_j(k).\partial W_{ji}(k)} \quad \text{and}$$

$$\frac{\partial J(k)}{\partial y_j(k)} = -e_j(k) \qquad (6)$$

The ordered partial derivatives of neural network outputs with respect to any given weight are, with L representing the number of delays at the neural network outputs to which feedback is applied [12]: where

$$\frac{\partial^+ y_j(k)}{\partial W_{ji}(k)} = \sum_{j=1}^{m} \Big(\frac{\partial y_j(k)}{\partial W_{ji}(k)} +$$

$$\sum_{q=1}^{L} \mu^q \frac{\partial y_j(k)}{\partial y_j(k-q)} . \frac{\partial^+ y_j(k-q)}{\partial W_{ji}(k)}\Big)$$

$$\frac{\partial^+ y_j(-1)}{\partial W_{ji}(k)} = \ldots\ldots . \frac{\partial^+ y_j(-L)}{\partial W_{ji}(k)} = 0 \qquad (7)$$

It is seen from this equation that ordered partial derivatives are computed sequentially. The first term on the right indicates the the static back-propagation algorithm, and the second term shows the effect of fed-back neural network inputs on weights. μ is a constant between 0 and 1, used to reduce the influence of previous inputs on weights since weights are renewed at every sample.

4. FORWARD AND INVERSE MODELLING OF PMSM

In order to model a system, a model has to be chosen that best describes the input-output behavior of the system. Although an exact mathematical model of the system is not needed in modeling a system by neural networks, an approximate model is required to determine the rank of the system so that the number of inputs and outputs of the neural network may be identified. Memory Neuron Networks have been developed which do not require a system model and aim to model the system by using only instantaneous input and output information, but whether such neural networks can model all dynamical systems is still a subject for investigation [13].

In this section, the linearized model of PMSM will be derived, and the inputs and outputs of the multilayer neural network will be determined. A mathematical PMSM model with respect to rotor reference, assuming $v_d=0$, is given below [1,14].

$$\begin{bmatrix} \dot{i}_q \\ \dot{i}_d \\ \dot{w} \end{bmatrix} = \begin{bmatrix} -\frac{r}{L} & -w & -\frac{\lambda}{L} \\ w & -\frac{r}{L} & 0 \\ \frac{3p^2\lambda}{8J} & 0 & -\frac{B}{J} \end{bmatrix} \begin{bmatrix} i_q \\ i_d \\ w \end{bmatrix} +$$

$$\begin{bmatrix} \frac{2}{\pi L} & 0 \\ 0 & 0 \\ 0 & -\frac{p}{2J} \end{bmatrix} \begin{bmatrix} v \\ T_L \end{bmatrix} \quad . \qquad (8)$$

As is seen from the state equations, PMSM is a nonlinear system. These state equations are linearized around equilibrium points, and the nonlinear relationship between v, the voltage applied to the motor, and w, the rotor speed, is given in discrete time by Eqn. (9):

$$\omega(k) = f(\omega(k-1), \omega(k-2),$$
$$\omega(k-3), v(k), v(k-1)) \qquad (9)$$

The forward neural network model of the motor may be chosen in accordance with this equation as follows:

$$\omega_m(k) = f(\omega_m(k-1), \omega_m(k-2),$$
$$\omega_m(k-3), v(k), v(k-1)) \qquad (10)$$

This equation describes a recurrent neural network model. This structure, known as parallel modeling, has certain advantages such as preventing noise. However, a series-parallel modeling structure is preferred because it is not certain whether the recurrent neural network parameters will converge [6]. In series-parallel modeling, the system's previous inputs are taken instead of the previous neural network outputs as delayed inputs to the neural network. Hence, the following feed forward neural network model is obtained for the forward model of the motor:

$$\omega_m(k) = f(\omega(k-1), \omega(k-2), \omega(k-3), v(k), v(k-1)) \tag{11}$$

In addition, the dynamic back-propagation algorithm needed for training the recurrent neural network requires intense computations. Hence, if the desired output value for a neural network output is known, it is appropriate to remove feedback, obtain a feed forward neural network structure, and use the static back-propagation algorithm.

The neural network model necessary for inverse modeling of the motor can be chosen from the derived mathematical motor model as follows:

$$v(k) = g(\omega(k), \omega(k-1), \omega(k-2), \omega(k-3), v(k-1)) \tag{12}$$

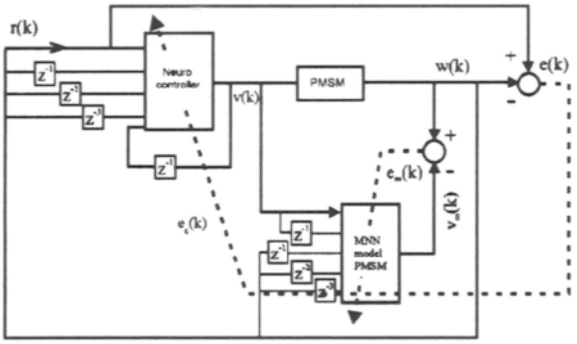


Figure 2. Indirect Adaptive Control of PMSM.

The adaptive control of PMSM using this feedback neural network structure is shown in Figure 2, with r(k) indicating reference rotor speed. In the inverse modeling of the motor, since the rotor speed in the k^{th} iteration is not known, r(k) can be taken as reference instead of w(k).

Because the desired output (rotor speed) is known for the forward neural network model of PMSMs, the error $e_m(k) = w(k) - w_m(k)$ needed for the training of the neural network is also known. However, because the desired output (voltage applied to the motor) for the inverse neural network model used as controller for the motor is not known, either the error $e_c(k)$ or the partial derivative of the system output, w(k), with respect to its input, v(k), has to be known in order to train the neurocontroller. Although an accurate mathematical model of the motor is not known, it can be determined using this derivative numerically $[\partial w(k)/\partial v(k)]$, and direct adaptive control of the motor can be realized without using its forward neural network model. This, however, is not the preferred method. It is more appropriate to train the neurocontroller by reflecting the motor reference speed tracking error, $e(k) = r(k) - w(k)$, backwards over the trained forward neural network model. Accordingly, the tracking error backpropagated onto the controller output can be determined using Eqns. (1) and (5).

$$e_c(k) = e(k).\frac{\partial w_m(k)}{\partial v(k)} \tag{13}$$

where W^1_{iv} indicates the hidden layer weights dependent on input v(k).

Self-controlled PMSM is driven by an inverter switched according to rotor position. Two methods can be used for speed control of PMSM by changing motor voltage. Motor is operated with the six step inverter and DC voltage of the inverter - shown v in Eqn[8]- is adjusted. MNN controller is trained to produce this voltage. As a second method, motor voltage is adjusted with PWM and controller MNN is trained to produce reference PWM voltage. The reference PWM voltage produced by MNN is converted to stator reference voltages - needed for sinusoidal PWM - by Eqn[14],

$$\begin{aligned} v^*_a &= v_r Cos(\theta) \\ v^*_b &= v_r Cos(\theta - 2\pi/3) \\ v^*_c &= v_r Cos(\theta + 2\pi/3) \end{aligned} \tag{14}$$

5. SIMULATION RESULTS

The modeling and indirect adaptive control of a small permanent magnet synchronous motor using a multilayer neural network has been realized in

the simulation study. The parameters of the PMSM used are :
R=3.5 ohm L=0.015 H P=8
J= 0.001 kg-m^2 B=0.00015 N-m/rad./sn.
λ=0.007 v-rad/sec v=10 volt-DC.

The simulation study was performed in two stages. First, the motor was modeled using neural networks, and the points to be emphasized in modeling were identified by inspecting the capability for generalization of the neural network. Next, adaptive control of the motor was realized using a trained neural network model. There is no method that gives a precise value for the number of processing elements in the hidden layer of the MNN, and this is generally found by trial and error. Simulation studies have shown that an MNN with 4 or 5 processing elements in the medium layer is sufficient for both neural network models.

Figure 3 shows the approximation performance of the motor's neural network model to motor speed, after the neural network has been trained for 800 seconds (200 000 samples) using the series-parallel method of motor modeling and the static back-propagation algorithm. An important point in system modeling is that the input applied to the system should be persistently excitation to bring out the dynamic behavior of the system completely. For this reason, the system was modeled using a sinusoidal input signal v(k)=10*sin(0.2*p*t(k)). In modeling, the load torque was changed randomly and, in Figure 3, it can be seen where the load torque is applied. During the first 5 seconds and last 10 seconds, the training of the neural network was stopped and the degree to which the model approximates the actual motor was determined. From the results, it may be emphasized that as the training period progresses, the neural network model exactly approximates the actual motor. Because convergence of the back-propagation algorithm is slow, however, and especially in the case of instantaneous sampling algorithms, the necessity of a large number of samples constitutes a serious disadvantage.

Using the forward MNN model, the inverse motor model—multilayer neurocontroller—was obtained, and indirect adaptive control of the motor was realized. The inverse motor model is trained during control of the motor by reflecting the tracking error over the forward motor model. If the forward motor model is not sufficiently accurate, therefore, convergence cannot be obtained for the MNN controller. It was found that the controller MNN coefficients did not converge unless the model MNN was trained at least once for every sample before the controller MNN was trained. Control of the motor using an MNN so as to track a reference speed of r(k) = 130*sin(0.15*p*t(k)) + 110*sin(6*p*t(k)) was realized accordingly, and the reference input tracking performance of the motor after 1500 seconds' training of the controller MNN is given in Figures 4.

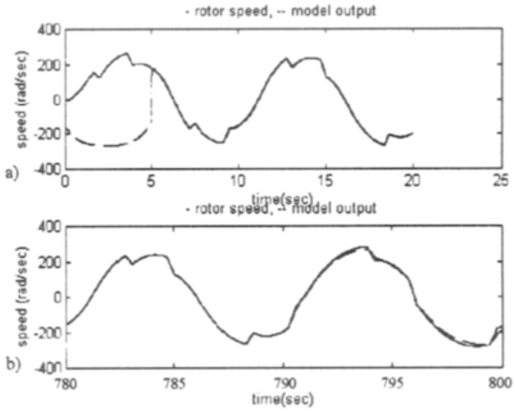


Figure 3. PMSM Modeling Performance of MNN. (Training of neural network was begun at the 5th second and was stopped at the 790th second.)

In Figure 4.a, during the last 10 seconds, the training of both the model MNN and the neurocontroller was discontinued. The voltage applied to the motor should not exceed its normal value during training of the neurocontroller. This constraint has been taken into account during the controller MNN's training, and in Figure 4.b, the controller MNN output and motor load torque for the last 20 seconds is shown. The generalization capability of the trained controller MNN for a different reference input is shown in Figure 4.c. The controller MNN output and load torque is shown in Figure 4.d. Figures 4 show that the MNN controller has an excellent control performance under the disturbance input.

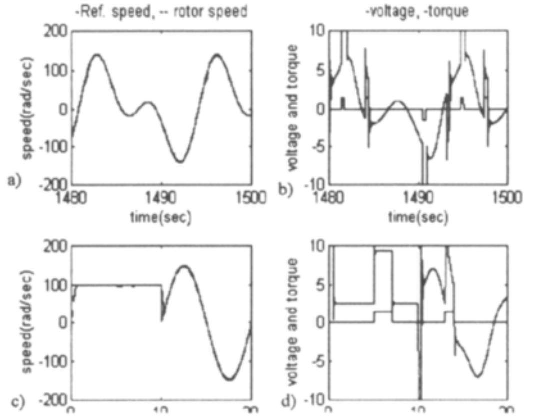


Figure 4. Reference Signal Tracking and Generalization Performance of Motor.

In Figure 4, PMSM was analyzed for fundamental component of six-step inverter output voltage and MNN controller was trained to produce inverter's DC supply voltage. The tracking performance of the PMSM supplied from PWM inverter is seen in Figures 5. MNN controller was trained to produce v_r and, reference stator voltages to be applied to motor were obtained using rotor position ,according to Eqn (14). In Figure 5.b, the phase a to neutral voltage of the PMSM, in Figure 5.c, actual q-axis voltage and, in Figure 5.d, voltage between phase a and negative terminal of DC supply is shown.

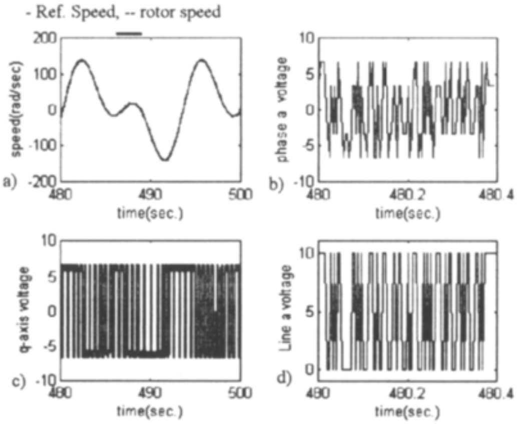


Figure 5. Tracking Performance of PWM Inverter-Fed Motor.

6. CONCLUSIONS

In this study, the capabilities of multilayer neural networks to model and adaptively control permanent magnet synchronous motors have been investigated. The sampled static back-propagation algorithm has been used for training the model MNN, and the sampled dynamic back-propagation algorithm for training the neurocontroller. In sampled back-propagation algorithms, a large number of samples are needed to completely train a neural network. On the other hand, adaptive controls are a necessity. Simulation results indicate that multilayer neural networks can be used as an effective method for the modeling and adaptive control of nonlinear systems.

7. REFERENCES

1. S. A. Nasar, I. Boldea and L. E. Unnewehr, Permanent Magnet, Reluctance and Self-Synchronous Motors, CRC Press, 1993.

2. M.A. El-Sharkawi and C.H. Huang, "Variable Structure Tracking of DC motor for High Performance Applications," *IEEE Transactions on Energy Conversion,* Vol. 4, No: 4, pp.643-650, 1989.

3. R.B. Sepe and J.H. Lang, "Real-Time Adaptive Control of Permanent Magnet Synchronous Motors," *IEEE Transactions on Industrial Applications,* Vol.27,No: 4, pp. 706-714, 1991.

4. K.R. Shouse and D.G. Taylor, "A Digital Self-Tuning Tracking Controller for Permanent Magnet Synchronous Motor," *IEEE Transactions on Control Systems Technology,* Vol. 2, No: 4, pp. 412-422, 1994.

5. E. Cerretu, A. Consoli and A.Raciti, "A Robust Adaptive Controller for Permanent Magnet Motor Drives in Robotic Applications", *IEEE Transactions on Power Electronics,* Vol. 10, No: 1, pp. 62-71, 1995.

6. K.S. Narendra and K. Parthasarthy, "Identification and Control of Dynamical System Using Neural Networks," IEEE Transactions on Neural Networks, Vol. 1, No: 1, pp. 4-27, 1990.

7. K.J. Hunt, D. Sbarbaro and R. Zbikowski, "Neural Networks for Control Systems - a Survey," *Automatica*, Vol. 28, No: 6, pp. 1083-1112, 1992.

8. J. Tanomaru and S.Omatu, "Process Control by On-Line Trained Neural Controllers," *IEEE Transactions on Industrial Electronics*, Vol. 39, No: 6, pp. 511-521, 1992.

9. S. Weerasooriya and M.A. El-Sharkawi, "Idenfification and Control of a DC Motor Using Back-Propagation Neural Networks," *IEEE Transactions on Energy Conversion*, Vol. 6, No: 4, pp. 663-669, 1991.

10. M.A. El-Sharkawi, A.A. El-Samahy and M.L. El-Sayed, "High Performance Drive of DC Brushless Motors Using Neural Networks," *IEEE Transactions on Energy Conversion*, Vol. 9, No: 2, pp. 317-322, 1994.

11. S. Haykin, *Neural Networks - A Comprehensive Foundation*, Macmillan College Publishing Company, Inc. US, 1994.

12. S.W. Piche, "Steepest Descent Algorithms for Neural Network Controllers and Filters, " *IEEE Transactions on Neural Networks*, Vol. 5, No: 2, pp. 198-212, 1994.

13. P.S. Sastry, G. Santharam and K.P. Unnikrishnan, "Memory Neuron Networks for Identification and Control of Dynamical Systems," *IEEE Transactions on Neural Networks*, Vol. 5, No: 2, pp. 306-319, 1994.

14. P.C. Krause., *Analysis of Electric Machinery*, McGraw-Hill Book Company 1986.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Robust Position Control for a DC Servomechanism Subject to Friction and Compliance

Yufeng Li, Bengt Eriksson and Jan Wikander

Department of Machine Design, Royal Institute of Technology, 100 44 Stockholm, Sweden

The position accuracy in a high performance motion control system is highly affected by vibrations due to compliance and nonlinear effects such as friction. This paper presents a nonlinear robust controller design which consists of three parts: vibration suppression through a disturbance observer, a PI velocity controller and a nonlinear position controller. Compared with a conventional linear cascaded controller with friction feedforward compensation, experiments show that with the proposed controller, not only is the positioning error reduced to within the range of the encoder resolution, it is also more robust to model uncertainties.

1. INTRODUCTION

Point-to-point positioning control is a frequently performed task in e.g., microelectronics manufacturing. One application is surface mounting robots (SMR) used to mount electronic components on circuit boards. High demands on the positioning accuracy result from the reduced size of modern electronic components and the high operational speeds required for achieving high productivity. In addition, robustness of the servo-system is also an important issue. Robustness must be ensured not only for system stability but also for performance. Differences in parameters among individual machines represent uncertainties due to nonlinear effects. It is required that the same controller settings can be used for all machines, i.e., without individual tuning.

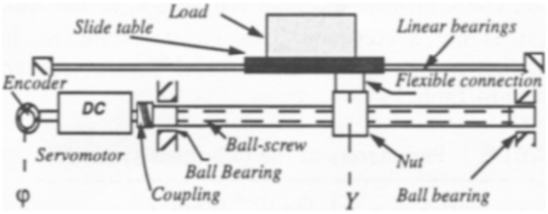


Fig. 1. The Y-axis of MYDATA surface mount robot

The experimental equipment used in this project is the Y-axis of the SMR, as shown in Fig. 1. The rotation of a current controlled DC motor is converted into a translational motion by a high precision ballscrew. A slide table attached to the ball-nut carries the loads at high velocities. The position is measured by an incremental encoder on the motor side, with a total moving range of 1.2 meters of the table. The goal of the control is to make the positioning as fast as possible with high end point positioning accuracy. However, some impediments make the goal difficult to achieve, i.e, the SMR has flexible elements and is subject to nonlinear friction. These characteristics greatly affect the accuracy of the position control as well as the achievable closed-loop bandwidth.

Vibration due to the flexibility may be suppressed to some degree by methods in terms of a disturbance observer [1]. Nonlinear friction which can not be handled by linear controllers is the main source of positioning inaccuracy. One common approach is to introduce a friction compensation based on a reasonably accurate friction model. The problem with this is that friction in general is dependent on many factors, for example: load, operating position, running time and machine wear. In fact, identification has shown that in the SMR, friction changes along the whole moving range. To keep the performance robust to the changing friction, either the friction model parameters or the control parameters must be updated. This requires more complicated control algorithms, e.g., on-line friction identification. There are many papers published on the friction problems and friction compensation [2].

The purpose of this paper is to present a design of a robust servo controller for the Y-axis of the SMR for positioning either at large or at small distances. Regardless of model uncertainties and nonlinear friction, the positioning error should be within the range of the measurement resolution. The following sections are organized as: section 2 presents the mathematical model of the compliant system and parameters identification. The design of the nonlinear robust position controller as well as vibration suppression with a disturbance observer are given in section 3. In section 4, experimental results are presented to examine the performance of the proposed controller, which is compared with a conventional linear controller. Conclusions are given in section 5.

2. SYSTEM IDENTIFICATION AND MODELLING

2.1 System modelling

One problem with the SMR is the vibrations caused by its compliant linkages between the motor and the load (the slide table with extenal load), which are the screw-shaft including the shaft coupling between the motor and the screw and the flexible connection between the nut and the slide, see Fig. 1. The vibrations may be provoked by a small torsion/bending of these compliant elements. Since it is not possible to have feedback on the load-end position, these vibrations may even cause instability in the robot.

Applying a chirp signal to the SMR (a sinusoidal signal with increased frequencies from 50~350 rad/s) and a fast Fourier transform (FFT) to the motor angular speed, the resulting power spectral density can be calculated, which shown in Fig. 2. It is found that high power density appears around the frequencies 150~300 rad/s. This implies that the most important resonant frequencies are within this range.

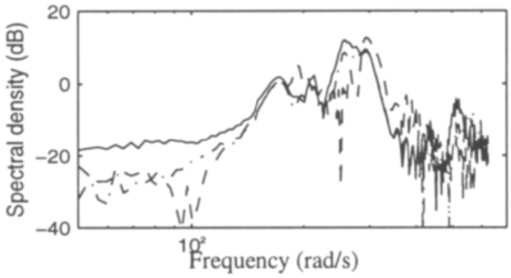


Fig. 2. Spectral analysis, slide table in the middle of the screw "solid", close to the motor end "dashed", away from the motor end "dash-dotted".

For simplicity, the experimental system is linearized around an operating point and modelled as two masses linked with a flexible shaft, as shown in Fig. 3. Neglecting the nonlinear friction, the transfer function from input current i_m to motor velocity v_m (translated motion on Y-axis from the motor angular velocity ω_m) is,

$$G_m(s) = \frac{v_m(s)}{i_m(s)} = \frac{(p/2\pi)K_m}{J_{m0}} \frac{s^2 + 2\xi_a\omega_a s + \omega_a^2}{s^3 + 2\xi_0\omega_0 s^2 + \omega_0^2 s} \qquad (1)$$

The resonant frequency ω_0 and anti-resonant frequency ω_a are,

$$\omega_0 = \sqrt{\frac{k_t}{J_l}(1 + J_l/J_{m0})}, \quad \omega_a = \frac{\omega_0}{\sqrt{1 + J_l/J_{m0}}} \qquad (2)$$

and the relative dampings ξ_0, ξ_a are

$$\xi_0 = \frac{d(1 + J_l/J_{m0})}{2\omega_0 J_l}, \quad \xi_a = \frac{\xi_0}{\sqrt{1 + J_l/J_{m0}}} \qquad (3)$$

where, K_m is the motor torque constant, p the pitch of the screw, J_{m0}, J_l the equivalent moment of inertia of the motor and the load, k_t the equivalent total stiffness of the linkages between the motor and load, d the damping coefficient and $\omega_a < \omega_0$, $\xi_a < \xi_0$.

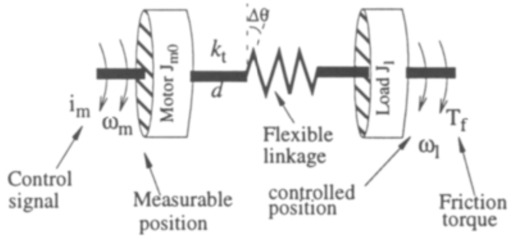


Fig. 3. Two- mass model of SMR Y-axis

Notice that the load is moving back and forth, i.e., the active length of the screw-shaft between the motor and the load changes during the operation. Since the shaft spring constant is a nonlinear function of its length and the length depends on the load position, consequently, the total stiffness k_t of the linkages is not a constant. This generates the model parameter uncertainty. The model parameters are given in Table 1 .

TABLE 1. Parameters of the two-mass system

Motor inertia: $J_{m0} = 1.38 \times 10^{-4}$ (kgm^2)

Load inertia: $J_l = 7.45 \times 10^{-4}$ (kgm^2)

Damping coefficient: $d = 0.02$ (Nm/rad/s)

Stiffness: $k_t = 5 \sim 10$ (Nm/rad)

Torque constant: $K_m = 0.356$ (Nm/A),

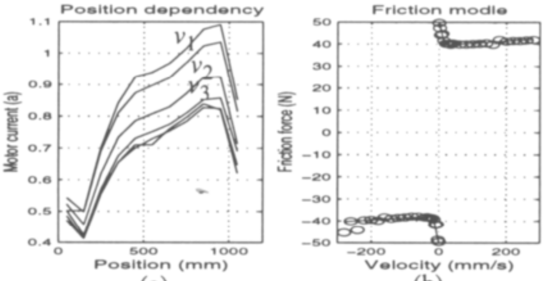
Screw pitch: p=0.040 m

2.2 Friction modelling

Several friction models have been proposed in the literature[2]. One often used model is,

$$F_f(v) = \left(F_c + (F_s - F_c)e^{-(|v|^{\sigma}/|v_s|)}\right)\mathrm{sgn}(v) + k_v v \quad (4)$$

where the friction force $F_f(v)$ is a function of velocity v. It is easy to verify that when the system is running at a constant velocity v_k, the relation, $F_f(v_k) = (p/2\pi)K_m i$, must be satisfied, hence the motor current i_m represents the level of the friction force. By repeating the experiments, $F_f(v)$ can be obtained. Fig. 4(a) shows that under different constant velocity (v_k, k = 1, 2,...), the measured motor current varies at different positions. It also shows that no simple function can be used to describe this friction dependence of the position. Therefore, the average value of the measurements is used for the friction estimation, $F_f(v)$ can be expressed by equation (4) , which is plotted in Fig. 4(b). The parameters of the model are given in Table 2 .



(a) (b)

Fig. 4. (a) Measured friction (b) Average value of friction "o" and friction model "-"

TABLE 2. Parameters of the friction model

Static: F_{s+} = 49.76 (N), F_{s-} = -49.66 (N)

Coulomb: F_{c+} = 39.36 (N), F_{c-} = -37.50 (N)

Viscous: $k_{v+} = k_{v-}$ = 0.01

Stribeck velocity: $v_{s+} = v_{s-}$ = 35 (mm/s)

σ=1.5

3. CONTROLLER DESIGN

3.1 Controller structure

Cascade control is often used in industry for motion control, the structure usually consists of an inner velocity loop and an outer position loop. This scheme is derived for the purpose of achieving good disturbance rejection. The shortcoming is that when the control servo is required to track reference trajectories with high values of speed and acceleration, the tracking capabilities of this scheme are unavoidably degraded. This can usually be solved by *decentralized feedforward compensation*, the reference signals: position y_d, velocity v_d, and acceleration a_d, are generated by a "trajectory planner". Then the "point-to-point" motion is controlled by decoupling the velocity control with v_d, the inertia force can also be fed forward by using a_d. Considering constraints on the maximum velocity and acceleration of the load, triangular or trapezoids velocity profiles are usually used since they have the shortest arrival time under a certain acceleration, e.g., with the maximum constant acceleration in the start phase and the maximum constant deceleration in the arrival phase. In many cases, a PI controller is enough for $G_v(s)$ to track the reference velocity and G_p can be just a proportional gain for a linear controller, see Fig. 5. The feedforward friction compensation in the figure uses the friction model specified in section 4.1. The disturbance observer is outlined in section 3.2.

Since the motion in the SMR is "point-to-point" positioning, only the final position is considered and the path of the motion is thereby not very important. Therefore, instead of using the piecewise linear time function to plan the trajectory, the desired velocity can be generated according to the position error.

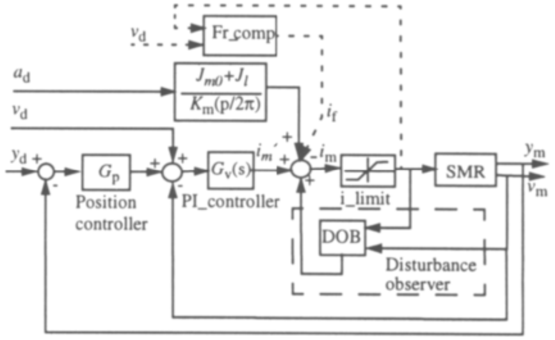


Fig. 5. Cascade control with decentralized feedforward compensation

Based on this idea, replacing the reference trajectory y_d with the target position, the reference velocity is generated by $G_p = v[\varepsilon(t)]$, which is a nonlinear function of the "the distance to the target" [3]. Now v_d is zero and a_d can be obtained from the derivation of $v[\varepsilon(t)]$. Here, G_p can be regarded either as a reference velocity generator, or as an adaptive position feedback gain according to the position error. $\varepsilon(t) = y_d - y_m(t)$ is the difference between the target position y_d and the actual position $y_m(t)$. The function $v[\varepsilon(t)]$ is defined as:

$$v[\varepsilon(t)] = v_0\frac{\varepsilon(t)/\varepsilon_0}{1+|\varepsilon(t)|/\varepsilon_0} \tag{5}$$

where, v_0 is a velocity limitation for a certain positioning distance and ε_0 is a positive real value. The advantage of using this function is that it is easy to take into account the constraints on the maximum torque and velocity related to the actuator limitations [3].

However, in equation (5) , the reference velocity at starting point, t=0, is a step function, which implies an infinite acceleration at t=0, i.e., $\left|\frac{d}{dt}v[\varepsilon(0)]\right| = \infty$. This should be avoided in SMR, i.e., the constraints (6) must always be satisfied,

$$v_0 \le v_{max}, \quad \left|\frac{d}{dt}v[\varepsilon(t)]\right| \le a_{max} \tag{6}$$

where, v_{max}, a_{max} are the maximum allowed velocity and acceleration. Modifying (6) by defining

$$v_0 = \begin{cases} sat(a_{max}t, v_{max}) & 0 \le y_m(t) < \Delta y/2 \\ sat(a_{max}t_0, v_{max}) & \Delta y/2 \le y_m(t) \le \Delta y \end{cases} \tag{7}$$

where,

$$sat(x, X) = \begin{cases} x & 0 \le x \le X \\ X & x > X \end{cases} \tag{8}$$

and $\Delta y = |y_d - y_m(0)|$ is the absolute value of the desired positioning distance (the distance between the start point and the target point), t_0, the time when the actual position y_m arrives at the point $\Delta y/2$.

To satisfy the second constraint of (6) under the given profile of v_0, ε_0 must be lower bounded. It can be proved that, for any ε_0>0, when $t \le t_0$, $\left|\frac{d}{dt}v[\varepsilon(t)]\right| \le a_{max}$ is always satisfied, while for $t > t_0$, the maximum of $\left|\frac{d}{dt}v[\varepsilon(t)]\right|$ is found at $|\varepsilon(t)| = \varepsilon_0/2$, consequently,

$$\left|\frac{d}{dt}v(\varepsilon)\right|_{max} = \frac{4}{27}\frac{v_0^2}{\varepsilon_0} \le a_{max} \Rightarrow \varepsilon_0 \ge \frac{4}{27}\frac{v_0^2}{a_{max}} \tag{9}$$

Notice that decreasing ε_0 can reduce the positioning time, however, too small ε_0 may cause limit cycles around the target position. Therefore, under the condition (9) , ε_0 must be a trade-off in practice.

3.2 Vibration suppression

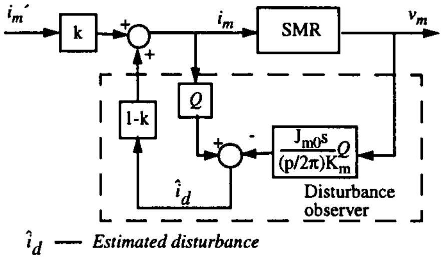

i_m'
k
+
+
i_m
SMR
v_m
Q
1-k
$\frac{J_{m0}s}{(p/2\pi)K_m}Q$
+
-
$\hat{i}_d$
Disturbance observer
$\hat{i}_d$ — Estimated disturbance

Fig. 6. Resonant ratio control

Vibrations caused by the compliant linkages of the SMR degrades the system performance. A disturbance observer based on *resonant ratio control* [1] and [4] is adopted in this paper to suppress the vibrations. The schematic expression is given in Fig. 6. Denote the original resonant ratio as H_0,

$$H_0 = \omega_0/\omega_a = \sqrt{1+R_0} \tag{10}$$

where, $R_0 = J_l/J_{m0}$, is inertia ratio. When R_0 is close to 1, i.e., ω_0 is close to ω_a, the flexible system is difficult to be controlled by high gain in the velocity loop due to the resonant peak.

The idea of resonant ratio control is to feed back (1-k) of the disturbance into the input signal. The virtual motor inertia can then be changed to $J_m = J_{m0}/k$, thus the new inertia ratio is

$$R = J_l/(J_{m0}/k) = kR_0 \qquad k \ge 1 \tag{11}$$

and the new transfer function from input i_m' to output v_m becomes:

$$G_m'(s) = \frac{v_m(s)}{i_m'(s)} = \frac{(p/2\pi)K_m}{J_m s}\frac{s^2+2\xi_a\omega_a s+\omega_a^2}{s^2+2c\xi_0\omega_0 s+c\omega_0^2} \tag{12}$$

It is easy to derive that,

$$c = \frac{1+R_0+Q(R-R_0)}{1+R_0} \tag{13}$$

where, $Q(s) = 1/(T_q s + 1)$ is a low pass filter, which satisfies $0 \le Q \le 1$ in all frequencies, thus, $c \ge 1$. Comparing equation (12) with equation (1), the relative damping and the resonant frequency are both increased by the factor $\sqrt{c}$.

The principle of choosing Q is that its cutoff frequency must be higher than the resonant frequency so that the torsional torque can be observed, but at higher frequencies, Q must be close to 0 in order to reject sensor noises. An optimal design of k and T_q is discussed in [4]. The frequency responses of G_m(s) and G_m'(s) are shown in Fig. 7(a), and the experimental verification of the effect of the disturbance observer can be seen in Fig. 7(b).

The disturbance observer has also good rejection for other low frequency disturbances.

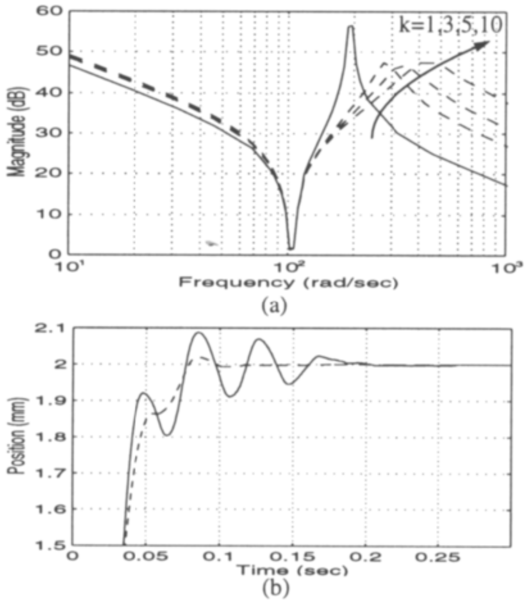


Fig. 7. The effect of the disturbance observer, G_m(s) (solid line) and G_m'(s) (dashed line); (a) The frequency analysis for k=1,3,5,10. (b) The time responses of a 2 mm positioning with k=3.

4. COMPARATIVE EXPERIMENTS

4.1 Controller parameters

The choice of the parameters of the PI controller $G_v(s) = k_p + k_i/s$ in the velocity loop is crucial to achieve good performance. In general, the larger the proportional gain k_p, the faster the controller response. However, since the bandwidth of the closed-loop can not go beyond the lowest anti-resonant frequency, i.e., around 100 rad/sec, the bandwidth is set around 80 rad/sec. To evaluate the performance, comparative experiments are made between the proposed controller and a linear controller with friction feedforward compensation, as shown in Fig. 5. In the linear controller, the reference velocity is generated by a trajectory planner using a triangular profile. Remark: friction compensation is not used for the proposed nonlinear controller because it does not gain much, in fact, it reduces the performance in some cases. However the compensation is needed for the linear controller in order to reduce the steady-state errors. The parameters of the velocity controller and disturbance observer in both control schemes are the same. Denote the proposed controller as C_1, the linear controller as C_2 and the friction feedforward compensation as Fr_comp. Table 3 gives the controller parameters for each controller.

TABLE 3. Parameters in the controllers

	Nonlinear controller C_1	Linear controller C_2
DOB	$k = 3, T_q = 0.002$	$k = 3, T_q = 0.002$
G_v(s)	$k_p = 0.05, k_i = 10$	$k_p = 0.05, k_i = 10$
G_p(s)	$G_p = 0.40$	$G_p = v[\varepsilon(t)]$
Fr_comp	No	Yes

Note: The selection of ε_0 in $v[\varepsilon(t)]$ is according to,

$$\varepsilon_0 = \begin{cases} (0.2 - 0.3)\Delta y & v_0 < v_{max} \\ \dfrac{4}{27}\dfrac{v_{max}^2}{a_{max}} & a_{max} t_0 \ge v_{max} \end{cases}$$

where, $a_{max} = 6.25\text{m/s}^2$ and $v_{max} = 1.00\text{m/s}$, the friction model used in feedforward compensation for C_2 is

$$F_f(v, F_e) = \begin{cases} -\text{sgn}(v)|F_f(v)| & |v| \ge D_v \\ -\text{sgn}(F_e) max(|F_f(v)|, |F_e|) & |v| < D_v \end{cases}$$

where, $F_f(v)$ is the friction force given in equation (4) and F_e is the applied (extenal) motor force, $D_v = 1.5$ mm/s specifies a small neighbourhood of zero velocity.

4.2 Experimental results

The experiment is implemented on the Y-axis of the SMR, position is measured from the encoder at the motor end. The resolution of the encoder is 20,000 pulses per revolution, or 2μm per pulse in terms of the translation motion to the Y-axis. Velocity signal is obtained by the difference between two continuous position measurements. The sampling rate in all experiments is 0.4 ms.

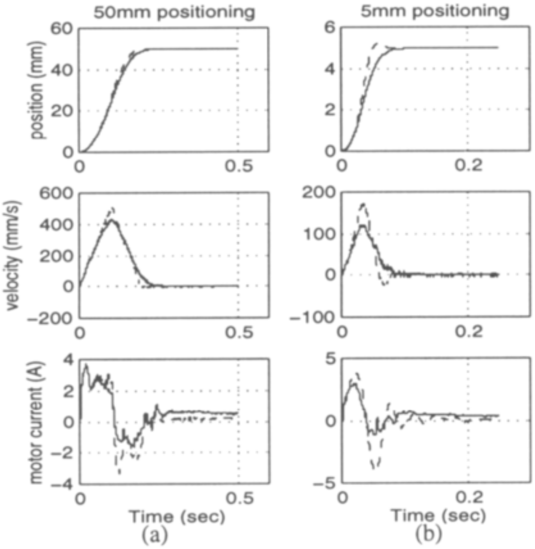


Fig. 8. The comparison of the two controllers, C_1 (solid) and C_2 (with fr_comp, dashed), (a) positioning 50mm, (b) positioning 5 mm.

TABLE 4. Positioning accuracy and time

Positioning distance	Accuracy and time	Controller C_1	C_2 (no Fr_comp)	C_2 (with Fr_comp)
50mm	ϕ (μm)	2μm	>15μm	8μm
	t_p (sec)	0.32s	0.32s	0.32s
5mm	ϕ (μm)	2μm	>10μm	4μm
	t_p (sec)	0.12s	0.21s	0.21s

The plots of two position responses, 50mm and 5mm are shown in Fig. 8. It can be seen that the response time of the controller C_1 is a little longer than that of C_2. This is due to the property of $v(\varepsilon)$. However, the setting time of C_1 is less than C_2 since C_1 has smoother velocity profile. Therefore, the total positioning time of the two controllers is about the same or even less with C_1 for small positioning distance. On the other hand, comparing the positioning accuracy, C_1 has much higher accuracy than C_2; actually, the accuracy of C_2 depends very much on the friction compensation. The experiments are repeated at different positions of the screw-shaft for testing the robustness, C_1 gives almost the same results, while C_2 does not because of the changing friction and the nonlinear stiffness. The accuracy of C_2 can not be guaranteed. Table 4 gives the numerical comparison results, in which, ϕ denotes the positioning accuracy and t_p denotes the total positioning time.

5. CONCLUSIONS

Nonlinearities usually present difficulties for motion controlled systems to achieve high performance control, since the nonlinearities can not be completely compensated by a linear controller.

This paper presents a nonlinear robust controller designed for a high accuracy positioning robot, SMR, which is subject to not only nonlinearities due to friction, but also the nonlinearity/uncertainty due to the system structure. The controller consists of three parts: vibration suppression, a PI-type velocity controller and a nonlinear position controller. Experiments show that the proposed controller has a strong robustness to the nonlinear friction. It guarantees the accuracy for both large and small distance positioning. Friction compensation is not needed because the nonlinear controller in the position loop compensates the remaining disturbance of the friction. In addition, it is more robust to parameter uncertainty.

6. ACKNOWLEDGMENT

This work was founded by Branschgroppen Datorstryd Mekanik and NUTEK. The experiment system was supplied by MYDATA AB.

REFERENCE

1. Y. Hori (1995). Vibration Suppression and Disturbance Rejection Control on Torsional Systems. *IFAC Motion Control,* pp 40-50, Munchen.
2. B. Armstrong-Helouvery, P. Dupont, C. Canudas de Wit (1994). A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction. *Automatica,* Vol. 30, NO.7, pp 1083-1183.
3. R. Gorez (1997), Sliding Mode As A Rational Approach to Robot Controller Design. *Proc. of the Workshop on Modelling and control of Mechanical Systems,* London.
4. Y. Hori, Y. Chun, H. Sawada (1996). Experimental evaluation of disturbance observer-based vibration suppression and disturbance rejection in torsional system, *Proc. of PEMC'96,* Budapest.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Mechanical object driven by the linear DC motor. Modeling, identification and control concepts

Jerzy Gustowski[a]

[a]Institute of Control and Computation Engineering, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warszawa, Poland
phone: (+4822)6607699, fax: (+4822)253719, e-mail: jgustowski@ia.pw.edu.pl.

Paper summarizes the author's efforts in the field of controlling the typical laboratory object - inverted pendulum, but driven by the nontypical electrical machine - linear DC motor. There are presented some results: algebraic model of the object, identification experiments conclusions, proposition of the control algorithm.

1. DESCRIPTION OF THE OBJECT

In the Institute of Control and Computation Engineering, Warsaw University of Technology, Poland, was constructed the laboratory rig which could be called *inverted pendulum* but it can be used also as a servomechanism or a model of an overhead crane.

Important and non-typical part of the rig is the linear DC motor (see Fig. 1) which is used to move directly (without any transmission gear) the movable cart to which the rotarily fixed arm is mounted. The motor consists of 1.5m. long stator to which two nonmovable coils are fixed. Between these coils the third one can freely move transporting the pendulum connected to it. The object is equipped with two optical increment encoders (linear and rotary) to measure the position of the cart and the angle of the arm, a precise sensor to measure the static driving force, a magnetic field meter to measure the electromagnetic induction in two slots of the linear motor. Linear DC motor is driven typically by the digital power amplifier fed by the TTL PWM pulse wave. The object is controlled by the industrial computer based on the VME bus architecture.

Motors typically used in similar applications are rotary ones. In our case, the reason of application of the linear motor was to make the way of driving the cart as natural as possible, especially to avoid using a transmission gear. The result of such solution, except the advantages mentioned, was a problem to precise the force driving the cart. This force is not directly proportional to the voltage or current signal which is an input of the motor but also depends to the position of the cart on the rail. This was the reason that the mathematical model of the linear drive was constructed. The model is based on the assumption that the linear motor may be considered as a magnetic long line. In this approach the most important thing is the determination of the functions describing the self-inductances of all the coils of the motor (L_1, L_2, L_t) and the mutual inductances (eg. L_{12}) particularly ones between the movable coil and nonmovable ones (L_{1t}, L_{2t}) (see Fig. 2.)

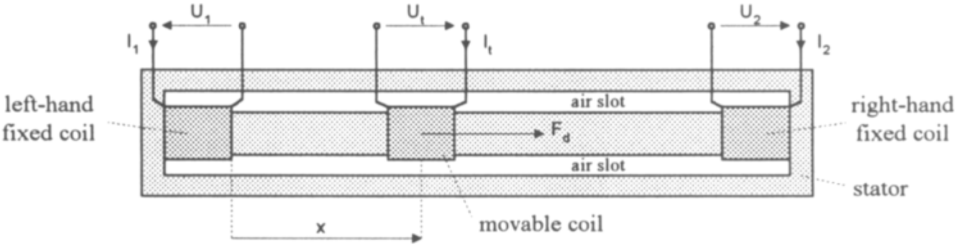


Figure 1. View of the DC motor

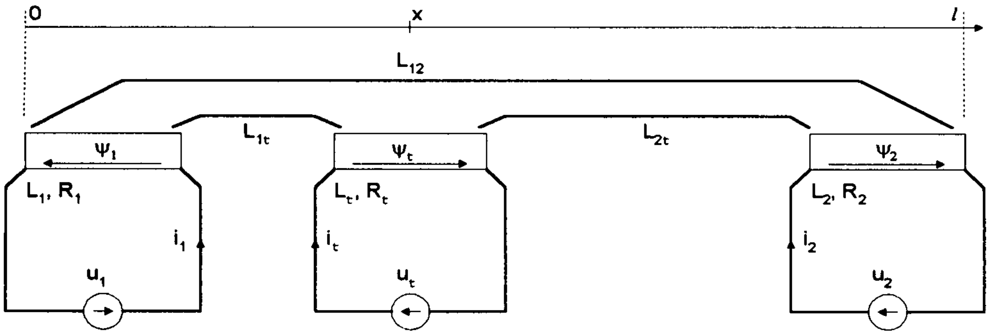


Figure 2. Scheme of electromagnetic connections in linear DC motor

2. MODEL OF THE OBJECT

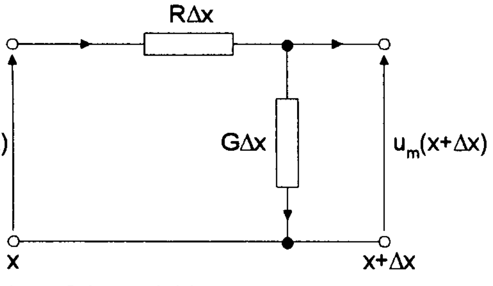
Model of the object can be derived from the Lagrange formalism [2]. To obtain the model it is important to determine dependencies of the associated magnetic fluxes ψ_1, ψ_2, ψ_t :

$$
\begin{aligned}
\psi_1 &= L_1 i_1 - L_{1t}(x)\, i_t - L_{12}\, i_2 \\
\psi_t &= -L_{1t}(x)\, i_1 + L_t(x)\, i_t + L_{2t}(x)\, i_2 \qquad (1) \\
\psi_2 &= -L_{12}\, i_1 + L_{2t}(x)\, i_t + L_2\, i_2
\end{aligned}
$$

and the values of all the inductances of the motor.

2.1. Linear motor as a magnetic long line

Assuming that the upper and the lower part of the motor are identical we treat the magnetic circuit of the stator as two magnetic long lines connected together parallelly. Elementary part of the magnetic long line of the length Δx could be shown as below:



Meaning of the variables:

ϕ - magnetic flux

u_m - magnetic voltage

R - elementary magnetic resistance

G - elementary magnetic conductance

Equations of the above four-terminal network are:

$$u_m(x) = \phi(x) R\, \Delta x + u_m(x + \Delta x)$$

$$\phi(x) = \phi(x + \Delta x) + u_m(x + \Delta x) G\, \Delta x$$

and after the transformation:

$$-\frac{u_m(x + \Delta x) - u_m(x)}{\Delta x} = R\, \phi(x)$$

$$-\frac{\phi(x + \Delta x) - \phi(x)}{\Delta x} = G\, u_m(x + \Delta x)$$

For $\Delta x \to 0$ we obtain:

$$-\frac{\partial u_m(x)}{\partial x} = R\, \phi(x) \qquad -\frac{\partial \phi(x)}{\partial x} = G\, u_m(x)$$

General solutions of above equations are:

$$\phi(x) = C_1\, sh\gamma\, x + C_2\, ch\gamma\, x$$

$$u_m(x) = -\frac{1}{Y}\left(C_1\, ch\gamma\, x + C_2\, sh\gamma\, x\right)$$

where: $\gamma = \sqrt{G R}$ $\quad Y = \sqrt{G/R}$

Linear motor treated as a magnetic long line can be presented as a circuit shown in Fig. 3. The Z_1 and Z_2 parameters are the magnetic resistances of the magnetic core closers of the both ends. The N_1, N_2 and N_t are the numbers of windings of each coil. All the parameters of so defined long line (Z_1, Z_2, R, G) can be evaluated knowing the geometry of the stator and magnetic permeabilities of its parts.

It is possible to determine the distribution of the magnetic flux produced by each coil separately. Assuming the linearity of the motor model the resultant distribution are the superposition of all the components. Evaluation of each magnetic flux distribution makes possible the determination of all the motor self- and mutual inductances.

Let's consider the left coil as a source of the magnetic flux.

$$i_t = i_2 = 0 \qquad \phi = \phi_1$$

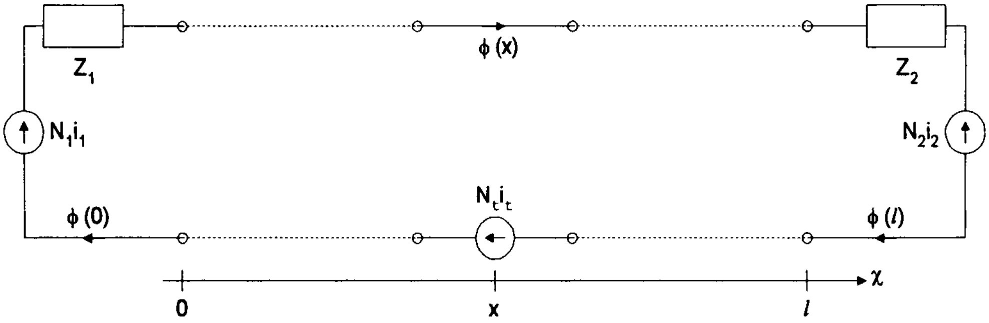

Z1
Z2
ϕ (x)
N1i1
N2i2
Ntit
ϕ (0)
ϕ (l)
0
x
l
χ

Figure 3. Scheme of the stator magnetic circuit

$$\phi_1(x) = \frac{ch\gamma\,(l-x) + Z_2Ysh\gamma\,(l-x)}{(Z_1+Z_2)Ych\gamma\, l + (1+Z_1Z_2Y^2)sh\gamma\, l} Y N_1 i_1$$

$$A = (Z_1+Z_2)Ych\gamma\, l + (1+Z_1Z_2Y^2)sh\gamma\, l$$

$$\phi_1(0) = (ch\gamma\, l + Z_2Ysh\gamma\, l)Y N_1 i_1/A$$

$$\phi_1(l) = Y N_1 i_1/A$$

$$L_1 = N_1\phi_1(0)/i_1 = N_1^2(ch\gamma\, l + Z_2Ysh\gamma\, l)Y/A$$

$$L_{12} = N_2\phi_1(l)/i_1 = N_1N_2Y/A$$

$$L_{1t}(x) = N_t\phi_1(x)/i_1 = N_tN_1\left[ch\gamma\,(l-x) + Z_2Ysh\gamma\,(l-x)\right]Y/A$$

Let's consider the right coil as a source of the magnetic flux.

$i_t = i_1 = 0 \qquad \phi = \phi_2$

$$\phi_2(x) = (ch\gamma\, x + Z_1Y\, sh\gamma\, x)Y N_2 i_2/A$$

$$\phi_2(0) = N_2 i_2 Y/A$$

$$\phi_2(l) = (ch\gamma\, l + Z_1Y\, sh\gamma\, l) N_2 i_2 Y/A$$

$$L_2 = N_2\phi_2(l)/i_2 = N_2^2(ch\gamma\, l + Z_1Y\, sh\gamma\, l)Y/A$$

$$L_{12} = N_1\phi_2(0)/i_2 = N_1N_2Y/A$$

$$L_{2t}(x) = N_t\phi_2(x)/i_2 = N_tN_2(ch\gamma\, x + Z_1Y\, sh\gamma\, x)Y/A$$

Let's consider the movable coil as a source of the magnetic flux (see Fig. 4).

$i_1 = i_2 = 0 \qquad \phi = \phi_t$

This is the case of two long lines serially connected. Let's consider step by step.

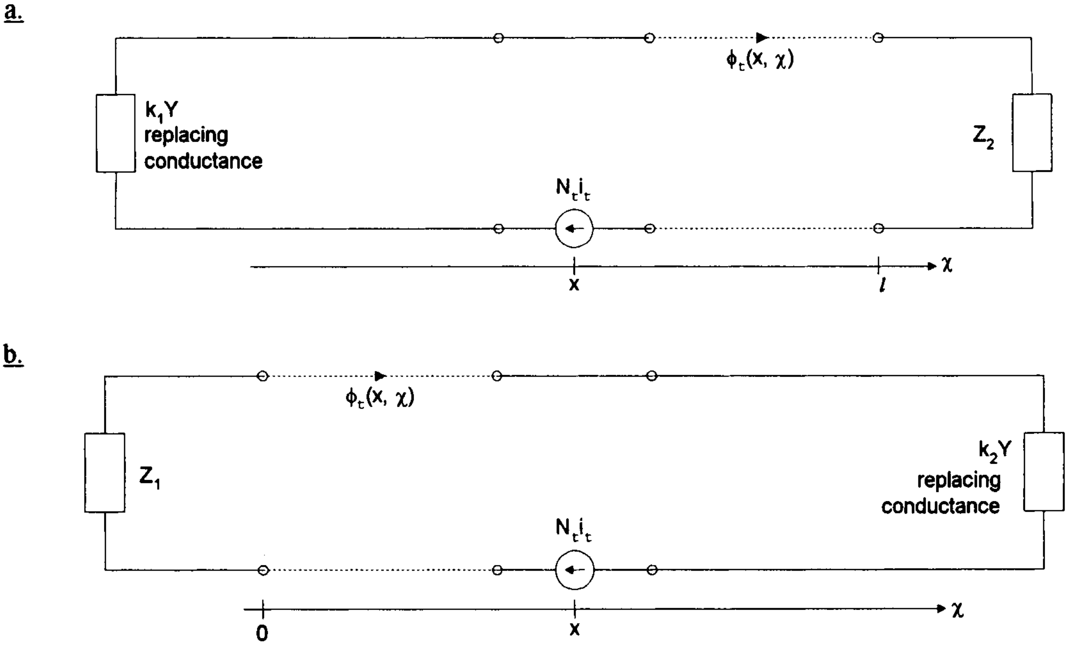

a.
k1Y
replacing
conductance
ϕt(x, χ)
Z2
Ntit
x
l
χ
b.
ϕt(x, χ)
Z1
k2Y
replacing
conductance
Ntit
0
x
χ

Figure 4. Two serially connected parts of the magnetic circuit supplied by the movable coil

From the Figure 4a:

$$k_1 = \frac{ch\gamma\, x + Z_1 Y\, sh\gamma\, x}{Z_1 Y\, ch\gamma\, x + sh\gamma\, x}$$

for $x \le \chi \le l$

$$\phi_t(x,\chi) = \frac{ch\gamma(l-\chi) + Z_2 Y sh\gamma(l-\chi)}{(k_1 Y + Z_2) Y ch\gamma(l-x) + (1 + Z_2 k_1 Y^3) sh\gamma(l-x)} Y N_t i_t$$

It is possible to verify that:

$$L_{2t}(x) = N_2\, \phi_t(x,l)/i_t = N_t\, \phi_2(x)/i_2$$

From the Figure 4b:

$$k_2 = \frac{ch\gamma\,(l-x) + Z_2 Y\, sh\gamma\,(l-x)}{Z_2 Y\, ch\gamma\,(l-x) + sh\gamma\,(l-x)}$$

for $0 \le \chi \le x$

$$\phi_t(x,\chi) = \frac{ch\gamma\,\chi + Z_1 Y sh\gamma\,\chi}{(Z_1 + k_2 Y) Y ch\gamma\, x + (1 + Z_1 k_2 Y^3) sh\gamma\, x} Y N_t i_t$$

It is possible to verify that:

$$L_{1t}(x) = N_1\, \phi_t(x,0)/i_t = N_t\, \phi_1(x)/i_1$$

Finally - the self-inductance of the movable coil:

$$L_t = N_t\, \phi_t(x,x)/i_t = (N_t^2 Y/2D) \cdot$$
$$\cdot \left[(1 + Z_1 Z_2 Y^2) ch\gamma\, l + (Z_1 + Z_2) Y\, sh\gamma\, l\right] +$$
$$+ \left(N_t^2 Y \big/ 2D\right) \begin{bmatrix} (1 - Z_1 Z_2 Y^2)\, ch\gamma\,(2x - l) + \\ + (Z_1 - Z_2) Y\, sh\gamma\,(2x - l) \end{bmatrix}$$

which means:

$$L_t = L_{t0} + \Delta L_t(x)$$

2. 2. Final version of the model equations

movable coil circuit

$$\left[L_{t0} + \Delta L_t(x)\right] \frac{di_t}{dt} - L_{1t}(x)\frac{di_1}{dt} + L_{2t}(x)\frac{di_2}{dt} +$$
$$+ \left[\frac{d\Delta L_t(x)}{dx} i_t - \frac{dL_{1t}(x)}{dx} i_1 + \frac{dL_{2t}(x)}{dx} i_2\right] \frac{dx}{dt} + R_t i_t = u_t(t)$$

left-hand coil circuit

$$L_1 \frac{di_1}{dt} - L_{12}\frac{di_2}{dt} - L_{1t}(x)\frac{di_t}{dt} +$$
$$- \frac{dL_{1t}(x)}{dx}\frac{dx}{dt} i_t + R_1 i_1 = u_1$$

right-hand coil circuit

$$L_2 \frac{di_2}{dt} - L_{12}\frac{di_1}{dt} + L_{2t}(x)\frac{di_t}{dt} +$$
$$- \frac{dL_{2t}(x)}{dx}\frac{dx}{dt} i_t + R_2 i_2 = u_2$$

movable mechanical part

$$(M + m)\frac{d^2x}{dt^2} + mr\frac{d^2\varphi}{dt^2}\cos\varphi - mr\left(\frac{d\varphi}{dt}\right)^2 \sin\varphi +$$
$$- \frac{i_t^2}{2}\frac{d\left[\Delta L_t(x)\right]}{dx} - i_t i_1 \frac{dL_{1t}(x)}{dx} + i_t i_2 \frac{dL_{2t}(x)}{dx} +$$
$$+ D\frac{dx}{dt} = F_h$$

rotary mechanical part

$$J\frac{d^2\varphi}{dt^2} + m\, r\frac{d^2x}{dt^2}\cos\varphi - m\, g\, r \sin\varphi + b\frac{d\varphi}{dt} = 0$$

3. DRIVING FORCE FUNCTION DETERMINATION

The evaluated mutual inductances depend of the position of the cart (movable coil) on the rail. This fact causes that the electrodynamic force driving the cart distinctly depends on the x coordinate. The value of the force vector in any point of the rail depends also on its sense, because the force function is asymmetric in relation to x (Fig. 5.)

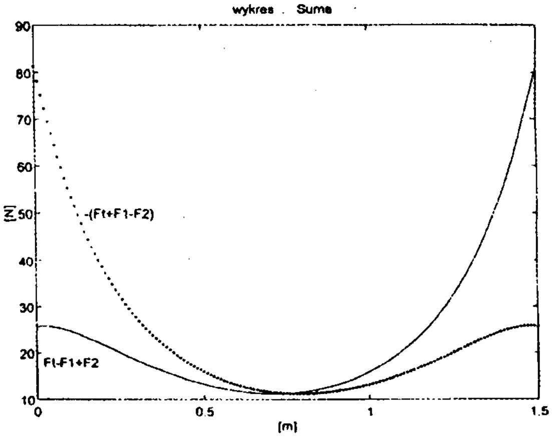


Figure 5. Functions of the driving forces depending of their senses

The values of the driving force obtained from the model are not the same as measured ones. Results of such comparison in the case of the static force (the cart is fixed to the force sensor) is given in the Fig. 6.

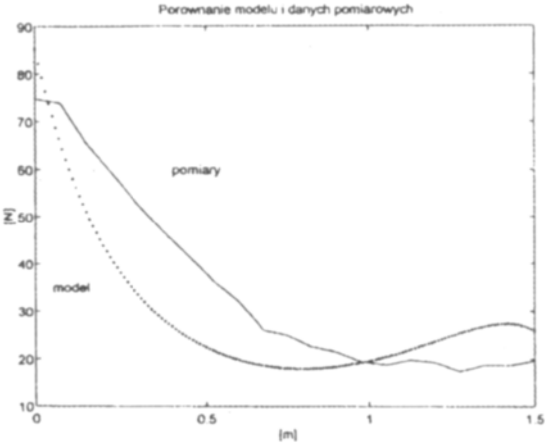


Figure 6. Comparison of the measured static force to one obtained from the model

The main control problems were not only the differences between the force theoretical and the measured one. It was very important to determine the dependency between the value of the PWM signal generated by the computer and the real value of the driving force in each position of the movable coil. This dependency were experimentally derived due to the multiple measurements. Results are presented in the Figure 7.

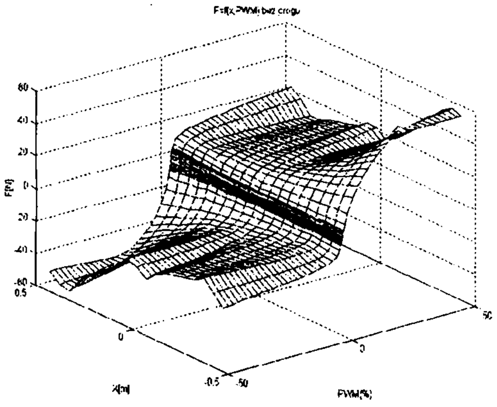


Figure 7. Driving force as a function of the PWM controlling signal and the position of the cart

More interesting is the inversion of the function shown in the Figure 7. It means that in real control one must know what PWM signal must by applied to the object in current position to obtain the value of the force required by the control algorithm. This function can be approximated by the formula:

$$PWM(F, x) = A(x)F + BF^5$$

where $A(x)$ has the form of the trinomial square of the variable x and B is a constant. This equation appeared to be good enough to apply the right control.

4. CONTROL ALGORITHM

The algorithm applied to the described object is based on the algorithm presented in [1]. This is the rule-based algorithm which is used to stabilize the inverted pendulum starting from its normal, stable position - hanging down. The algorithm has two phases: first - swinging the pendulum to increase its energy to be enough to raise to the upper vertical position and then to stabilize it in this position.

Rules used in the algorithm:

1. swinging:

$u_1 = F_{max}\, sign[\omega \cdot (|\varphi| - \pi/2)]$

2. soft landing in the stabilization zone:

$u_2 = u_1 \{[0.5 \cdot (sign(\varphi \cdot \omega) + 1)] + [0.5 \cdot (sign(E) + 1)] \cdot [0.5 \cdot (1 - sign(\varphi \cdot \omega))]\}$

where: $E = m\, r\, g \cdot (1 - \cos\varphi) + f_p \cdot \varphi \cdot \omega - 1/2 \cdot J\omega^2$

3. stabilization:

$u_3 = min(F_{max}, k_1 \cdot x + k_2 \cdot \varphi + k_3 \cdot v + k_4 \cdot \omega) \cdot \{0.5 \cdot [sign(|\varphi| - \varepsilon) - 1]\} + u_2 \cdot \{0.5 \cdot [sign(|\varphi| - \varepsilon) + 1]\}$

5. rail limitations:

$F = u_4 = u_3 \cdot \{0.5 \cdot [sign(x_g - |x|) + 1]\} + F_{max} \cdot sign(x) \cdot \{0.5 \cdot [sign(x_g - |x|) - 1]\}$

where:

φ - angle of the pendulum arm
ω - angular speed of the arm
x - position on the rail

The general idea of the algorithm is to swing the arm by moving the cart so long and so strong as the arm has enough energy to go without any force to the upper vertical position and then nearly this position to switch the control strategy from swinging to stabilization. In the stabilization zone typical PD algorithm is applied.

Two modifications of the presented algorithms are proposed. The first proposition is to divide the rotation space of the pendulum arm into three subspaces not two as above. This is proposed to avoid as long as possible the situation in which the arm during the stabilization phase swings a bit too much then the ε angular limit and causes the switching on the swinging phase. Such algorithm is very flexible to the choice of the value of ε. In the proposed algorithm in the narrowest subspace the classical PD controller is applied. In the $\varepsilon \leq /\varphi/ \leq \delta$ subspace the force pushing back the arm to the PD subspace is produced. In the outer subspace the swinging algorithm is used.

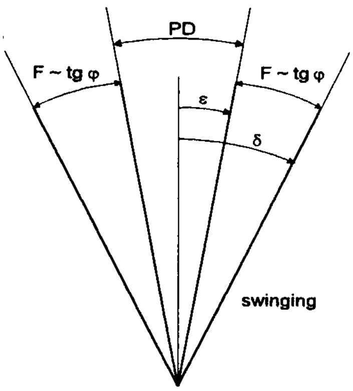


Figure 8. Subspaces in arm rotation space

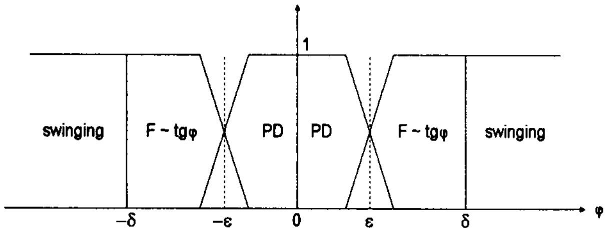


Figure 9. Policy of switching of the control strategies

The force proportional to the $tg\varphi$ is the force which is able to hold the arm in the stable inclination of $\varphi \neq 0$ by moving the cart to this direction to which the arm is going to fall.

To make the transition between the swinging and the stabilization phases more soft it is proposed to fuzzyfy the border for the $\varphi = /\varepsilon/$ (see Fig. 9).

In the algorithm presented in [1] rule 2 describes the decision when to switch off the force used to swing the arm and to let it freely go to the upper vertical position. It is done by evaluating the arm energy which could be not very precise. It is proposed to swing the arm in such a way that it could move to the upper position in the same way (it means the correlation between its momentary angle and speed) as it fall freely down which could be previously registered without any problems.

Results of application described algorithm to the inverted pendulum driven by linear DC motor are shown in Figure 10.

REFERENCES

1. A. Turnau, Fuzzy and Rule-Based Controller for Cart-Pole System on Finite Rail, Proceedings of the Conference on Modelling and Simulation - ESM 94, Barcelona, Spain, Jun. 1-3, 1994
2. D. C. White, H. H. Woodson, Electromechanical Energy Conversion, John Wiley & Sons, Inc., New York, 1959

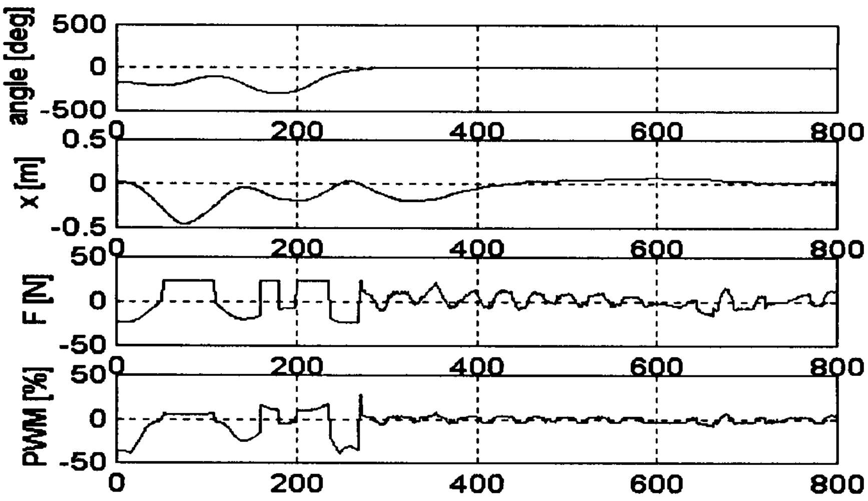


Figure 10. Controlling of the inverted pendulum

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Determination of the Transfer Function for the Mechanical Objects Along Transitional Response

K. Sarkauskas*, R. Dapkus* and M. Rackaitis**

*Faculty of Electrical Engineering and Control Systems, Kaunas University of Technology,
Studentu Str. 48 - 314, Kaunas LT-3028, LITHUANIA
** Research Center "Vibrotechnika", Kaunas University of Technology,
Studentu Str. 65, Kaunas LT-3031, LITHUANIA

The paper deals with the mechanical object's transfer function determination using theoretical presumption, which earlier practically was not possible apply because of large error of the calculations. The determination of transfer function could be carried out along transitional response of the mechanical object by implementation of the proposed new mathematical method and using new created software.

1. INTRODUCTION

Mechanical object in dynamics can be described by differential equations. They could be determined by analytical method taking into account the principle of systems operation, properties of used materials and general construction of the mechanical system. But in most cases mathematical description of real mechanical system differs from analytical model because of not estimated some parameters or technological deflection and used material's properties variation. In order to eliminate these inadequacy the correction and reserve coefficients are used.

In this paper is proposed a new mathematical method which allows to make more accurate previously defined or to obtain new exact differential equations of real mechanical system by experimental data.

The solution of differential equations with nonvariable coefficients and zero initial conditions, exist and is unique and only one differential equation's answer for any graphic or tabular given data. Finally the solutions in this direction is approximate and with the same data we obtain different transfer functions. The errors of approximation can be minimal if the calculations will be done with computer; and finally - the solution can be the only one. The transfer function detection according the experimental data is the most efficient use of the "area" method [1, 2].

The expression of transfer function is shown as

$$W(p) = K\frac{\sum_{j=1}^{n} b_j p^j + 1}{\sum_{i=1}^{m} a_i p^i + 1}, \tag{1}$$

and only in exceptional case could be

$$W(p) = \frac{K}{1+\sum_{i=1}^{m} a_i p^i} . \tag{2}$$

The solution of transient response ***F(t)*** can be determined with Laplace transformation

$$F(p) = \frac{1}{p}W(p) . \tag{3}$$

The first area S_1 can be determined (as shown in Fig.1)

$$S_1 = \int_0^\infty \left[K - F(t)\right]dt = \lim \int_0^\infty e^{-pt}\left[K - F(t)\right]dt =$$

$$= \lim_{p\to\infty} \pounds\left\{K - F(t)\right\}, \qquad (4)$$

where $\pounds\left\{K - F\right\}$ - Laplace transformation of function ***(K-F(t))***; then

$$S_1 = \lim_{p\to 0}\left[\frac{k}{p} - F(p)\right] = \lim_{p\to 0}\left[\frac{1}{p}(K - W(p))\right] =$$

$$= \lim_{p\to 0} K \frac{\sum_{i=1}^{\infty} a_i p^{i-1}}{1 + \sum_{i=1}^{\infty} a_i p^i} = Ka_1. \qquad (5)$$

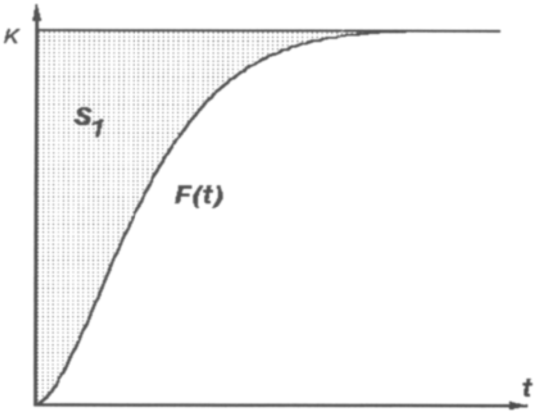


Figure 1. Determination of the area S_1

In the first approximation

$$W_1(p) = \frac{K}{a_1 p + 1}, \qquad (6)$$

and

$$F_1(p) = \frac{1}{a_1 p + 1}\frac{K}{p}. \qquad (7)$$

The area S_2 can be detected (as shown in Fig. 2).

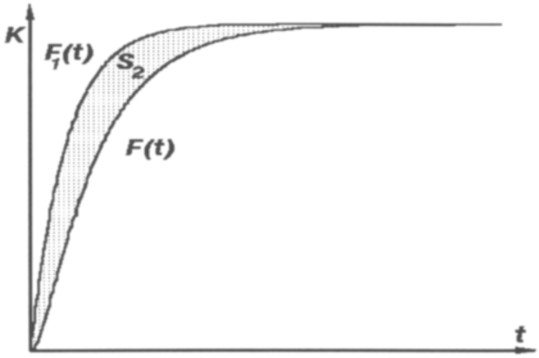


Figure 2. Determination of the area S_2

$$S_2 = \int_0^\infty\int_0^\infty \left[F_1(t) - F(t)\right]dtdt =$$

$$= \lim_{p\to 0}\frac{1}{p} L\left\{F_1(t) - F(t)\right\} =$$

$$= \lim_{p\to 0}\frac{1}{p}\left\{\frac{K}{a_1 p + 1} - \frac{K}{1 + \sum_{i=1}^{\infty} a_i p^i}\right\} =$$

$$= \lim_{p\to 0}\frac{1}{p}\left\{\frac{\sum_{i=1}^{\infty} a_i p^i - a_1 p}{(a_1 p + 1)(1 + \sum_{i=1}^{\infty} a_i p^i)}\right\} =$$

$$= \lim_{p\to 0}\frac{K\sum_{i=2}^{\infty} a_i p^{i-1}}{(a_1 p + 1)(1 + \sum_{i=1}^{\infty} a_i p^i)} = Ka_2\ . \qquad 8)$$

Further calculations of areas S_j are analogous.

The solution of problem has more difficulties if the appearance of the transfer function is like in formula (1). Then after areas calculation we will have

$$\begin{cases} S_1 = K(a_1 - b_1), \\ S_2 = K(a_2 - b_2 - b_1 S_1), \\ S_3 = K(a_3 - b_3 - b_1 S_2 - b_2 S_1), \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ S_j = K(a_j - b_j - \sum_{i=1}^{j-1} b_i S_{j-i}). \end{cases} \tag{9}$$

In order to determine coefficients *a* and *b* we must have *m+n* equations. If *j>n*, then equations become

$$S_j = -K\sum_{i=1}^{j-1} b_i S_{j-i}. \tag{10}$$

The theoretical solution of these equations is not complicated, but in reality very often if just one area has the negative value, the solutions of transfer function are not stable. Well known "area's" method is not suitable for immediate integration, so expression (4) should expanded into Taylor series e^{-pt}. This allows further "area" calculations. The precision of the "areas" method is limited and the form of transfer function could be determined only approximately.

2. DESCRIPTION OF THE METHOD

The authors of this paper have developed method, which has not the imperfections destribed above. There is presented new method for establishment the transfer function of control object along experimental data. The approaching method is modification of area method. The advantage of this method is that for calculations is used analytical expressions and its conclusions, but not approximation using Taylor series. The method could be implemented into practice only using PC, because of the area calculations must be carried out with high precision. The calculation is performed in relative units, bringing scale and dimension of transfer function after that.

The transitional response of transfer function could be distributed in three groups: functions with integration terms, inertial functions and functions with differential terms (Fig. 3) - they can represent majority of real systems.

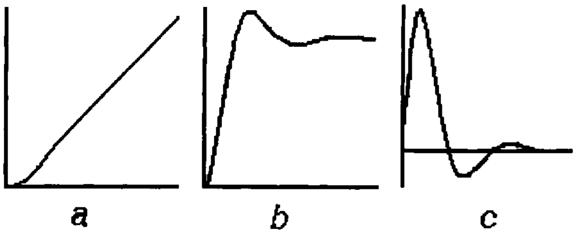


Figure 3. Family of transitional response

The family of this transitional response can be three times bigger if one estimate the functions with delay and the functions with oscillation. Existing "areas" method work only with transitional response, shown in figure 3b. Transfer function with integration (Fig 3a) or differentiation (Fig 3c) terms could be transferred into transitional response (Fig. 3b) after complementary differentiation or integration [3].

The transfer function, which transitional response represented in figure 3b, has form

$$W'_{(p)} = \frac{1}{p} W_{(p)}. \tag{11}$$

After differentiation

$$\frac{dF(t)}{dt} \div pW'(p) - f(0), \tag{12}$$

and if initial conditions are zero, then

$$\frac{dF(t)}{dt} \div W(p), \tag{13}$$

and the transitional response will have the shape as shown in figure 3b. This operation should be repeated as many times as needed and will be included in final representation of transfer function.

The transfer function, which transitional response is represent in figure 3c, has form

$$W'(p) = pW(p). \tag{14}$$

After integration we have

$$\int_0^t f(t)dt + \frac{W'(p)}{p} = W(p) \tag{15}$$

and the transitional response will have the same shape as it is shown in figure 3b. In this case further we examine only this transitional responses. In such systems the relation between n and m can be easy detected according functions and their derivative values:

$$\begin{aligned} &f(0)=0,\ f'(0)=0, && \text{where} && m<n,\\ &f(0)\neq 0,\ f'(0)=0, && \text{where} && m=n,\\ &f(0)\neq 0,\ f'(0)\neq 0, && \text{where} && m>n. \end{aligned} \tag{16}$$

The method was created for that casen of $n \geq m$. When $\varphi(t)$ is given graphically or by tabular transitional response in relative values, its Laplace transform will be

$$\varphi(p) = \frac{1}{p} W(p), \tag{17}$$

where

$$W(p) = \frac{\sum_{j=1}^{n} b_j p^j + 1}{\sum_{i=1}^{m} a_i p^i + 1}. \tag{18}$$

The integration error of transitional response with unit jump can be determined as marked area in figure 4, when $K=1$

$$S_1 = \int_0^\infty [1-\varphi(t)]dt. \tag{19}$$

The rules of integration declare

$$\int_0^\infty f(t)dt = \lim_{p\to 0} \frac{F(p)}{p} - f(0), \tag{20}$$

and integration value of first area will be

$$S_1 = \int_0^\infty \left[1-\varphi(t)\right]dt = \lim \int_0^\infty e^{-pt}\left[1-\varphi(t)\right]dt =$$

$$= \lim_{p\to\infty} L\left\{1-\varphi(t)\right\} \tag{21}$$

$$S_1 = \lim_{p\to 0}\left[\frac{1}{p} - \varphi(p)\right] = \lim_{p\to 0}\frac{1}{p}\left[1-W(p)\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\,\frac{1+\sum_{i=1}^{m} a_i p^i - 1 - \sum_{i=1}^{n} b_i p^i}{1+\sum_{i=1}^{m} a_i p^i} =$$

$$= \lim_{p\to 0}\frac{\sum_{i=1}^{m} a_i p^{i-1} - \sum_{i=1}^{n} b_i p^{i-1}}{1+\sum_{i=1}^{m} a_i p^i} = a_i - b_i \tag{22}$$

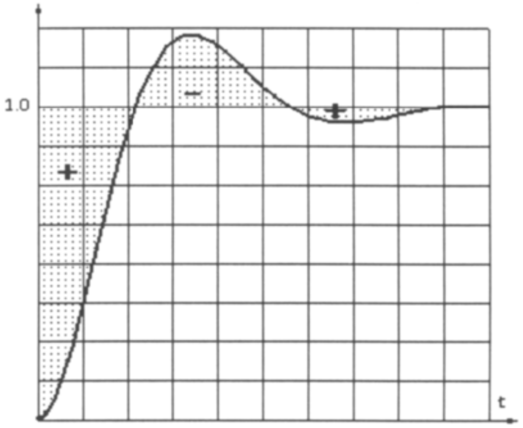


Figure 4. Determination of the area S_1

In the first approximation

$$W_1(p) = \frac{1}{S_1 p + 1}. \tag{23}$$

The integration error of transitional response of second area can be detected as marked area in figure 5, when $K=1$. The integration value of second area will be

$$S_2 = \int_0^\infty\int_0^\infty [\varphi_1(t) - \varphi(t)]dtdt = \lim_{p\to 0}\frac{1}{p}\frac{1}{p}\left[(W_1(p) - W(p))\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\frac{1}{p}\left[\frac{1}{1+S_1p} - \frac{1+\sum_{i=1}^{n} b_i p^i}{1+\sum_{i=1}^{m} a_i p^i}\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\frac{1}{p}\left[\frac{1+\sum_{i=1}^{m} a_i p^i - (1+S_1p)(1+\sum_{i=1}^{n} b_i p^i)}{(1+S_1p)(1+\sum_{i=1}^{m} a_i p^i)}\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\frac{1}{p}\left[\frac{1+\sum_{i=1}^{m} a_i p^i - 1 - S_1p - \sum_{i=1}^{n} b_i p^i - S_1p\sum_{i=1}^{n} b_i p^i}{(1+S_1p)(1+\sum_{i=1}^{m} a_i p^i)}\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\left[\frac{\sum_{i=1}^{m} a_i p^{i-1} - S_1 - \sum_{i=1}^{n} b_i p^{i-1} - S_1\sum_{i=1}^{n} b_i p^i}{(1+S_1p)(1+\sum_{i=1}^{m} a_i p^i)}\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\left[\frac{\sum_{i=1}^{m} a_i p^{i-1} - a_1 + b_1 - \sum_{i=1}^{n} b_i p^{i-1} - S_1\sum_{i=1}^{n} b_i p^i}{(1+S_1p)(1+\sum_{i=1}^{m} a_i p^i)}\right] =$$

$$= \lim_{p\to 0}\frac{1}{p}\left[\frac{\sum_{i=2}^{m} a_i p^{i-1} - \sum_{i=2}^{n} b_i p^{i-1} - S_1\sum_{i=1}^{n} b_i p^i}{(1+S_1p)(1+\sum_{i=1}^{m} a_i p^i)}\right] =$$

$$= \lim_{p\to 0}\frac{\sum_{i=2}^{m} a_i p^{i-2} - \sum_{i=2}^{n} b_i p^{i-2} - S_1\sum_{i=1}^{m} b_i p^{i-1}}{1+\sum_{i=1}^{m} a_i p^i} =$$

$$= a_2 - b_2 - S_1 b_1. \qquad (24)$$

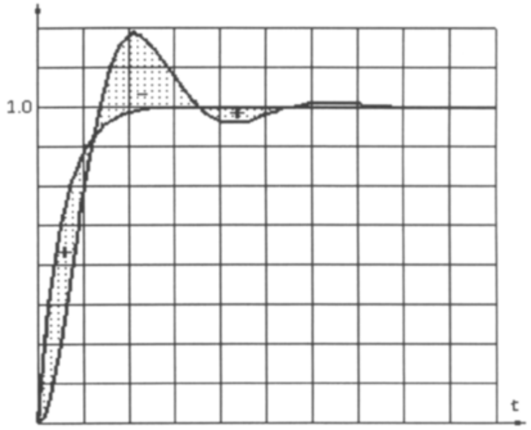


Figure 5. Determination of the area S_2

The integration value of third area will be

$$S_3 = a_3 - b_3 - b_1S_2 - b_2S_1. \qquad (25)$$

and etc.

The integration value of k - area will be

$$S_k = a_k - b_k - \sum_{i=1}^{k-1} b_i S_{k-i}. \qquad (26)$$

In order to determine coefficients a and b we must have $m+n$ or more equations. The system of linear algebraic equations will be:

$$\begin{cases} a_1 - b_1 = S_1, \\ a_2 - b_2 - b_1S_1 = S_2, \\ a_3 - b_3 - b_2S_1 - b_1S_2 = S_3 \\ \dots\dots\dots\dots\dots\dots\dots \\ a_1 - b_1 - \sum_{i=1}^{k-1} b_i S_{k-i} = S_k, \\ \dots\dots\dots\dots\dots\dots\dots \\ -\sum_{i=1}^{j-1} b_i S_{j-i} = S_j \end{cases} \qquad (27)$$

To solve this matrix is not easy, because the number of unknown variables is bigger than the number of the equations. On the other side, when the "area" becomes negative, the further solution has no sense, because transfer functions become unstable.

3. CONCLUSION

According to the proposed method's theoretical presumptions the solution exists and is only one if the transfer functions of identification object have not the roots in the numerator.

The suggested method could be named as the consecutive approach method. The essence and realisation sequence of the method for transfer function determination could be described as follows. At the beginning it is necessary to find the first area, which make difference between unit influence and experimental dynamic characteristic of identification object. Obtained value is first polynomial coefficient of transfer function. The next step is to define the second area, which make difference between transitional response of first approximation of transfer function and experimental dynamic characteristic of identification object. Obtained value is second polynomial coefficient of transfer function. Further is necessary to determine successive areas, which make difference between transitional response of early approximation of transfer function and experimental dynamic characteristic of identification object. Obtained value is corresponding to polynomial coefficient of transfer function. The end of calculations is determined when the sign of the area become negative.

Other interesting question - practical usage of this method. Existed difficulty in the detection of areas: it was necessary to write the computer program for differential equation's solution and area calculation, further after each identification step it was needed rewrite program for differential equation's solution and area calculation, because the power of differential equation and area's integration cycle increase step-by-step. Now this important problem is possible to solve using new software CADoVS (Computer Aided Design of Various System), created in Control Engineering Department of Kaunas University of Technology (more information about this package one can find at http://ediaclit.vtk.ktu.lt/SFAsoft).

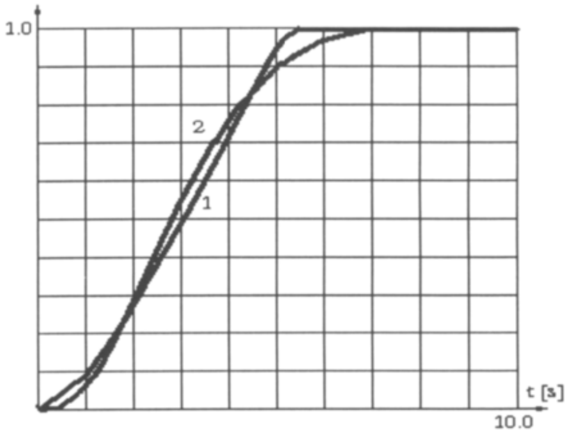


Figure 6. Transitional responses: 1 - experimental curve, 2 - Transitional response of determined transfer function

The significance of the proposed consecutive approach method could be illustrated in figure 6, where curve 2 obtained using this method highly correlates to the experimental curve 1.

REFERENCES

1. B.P. Welford, Note on a method for calculation corrected sums of squares and products, Technometrics, V. 4, No. 3 (1962).
2. B. Qvarnstrom, Transfer function determination in presence of noise for a set of significant input functions, Sweden Automatic Process Control, 1960.
3. K. Sarkauskas. Determination of dynamical characteristics of controlled objects from experimental date, *Elektrotechnics, Vilnius*, No.8, (1981), 131-143.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Smooth trajectory planning for planar laser cutting system

Miran Rodič, Aleš Hace, Martin Terbuc, and Karel Jezernik

University of Maribor, Faculty of Electrical Engineering and Computer Science, Institute for Robotics,
Smetanova 17, SI-2000 Maribor, Slovenia
phone: (++386) 62 220 7308 , fax: (++386) 62 220 73 15, e-mail: miran.rodic@uni-mb.si

Trajectory planning algorithms that are used in the 2DOF non-rigid laser cutting system are described in our paper. As the mechanism used in the application is not rigid, some restrictions are given for its operation, especially regarding trajectory planning. Thus changes of speed are an important issue. To preserve the smoothness of motion, an algorithm using $\sin^2$ acceleration profile was used. It had to be modified in order to avoid resonance frequencies and ensure timely decelerations in order to stop the machine without braking. Paper also describes the use of circular splines for the smoothing of sharp corners between following programmed trajectories.

1. INTRODUCTION

In recent years lasers were introduced into many different areas. They found their way into medicine, industry and even our homes. Besides measurements and transfer of informations also the cutting of materials can be performed using lasers. Combined with the modern computer technology using advanced algorithms, laser cutting systems became a growing area of industrial applications.

In our institute a laser cutting system for the small series production was developed. Currently it is used in the working process and is operating full time. Besides performance demands also a price of product was of great importance, since many laser cutting machines exist on the market.

The system enables laser cutting in the horizontal plane with the workspace of 1m x 2m using laser with the optical power of 500W. As a programming language standard G-code is used. Thus direct use of existing tools for conversion from CAD software (for example AutoCAD) to G-code is possible. Also no additional training is necessary for the CNC programmers, that have been previously trained on different machines using same program language.

During the operation velocity of movement is changing. In order to assure smooth trajectories, the velocity has to be continuos, which means that it has to change in the continuos manner. In our application algorithm using $\sin^2$ acceleration profile was used. Besides assuring of the smooth movement it is also implementing the avoidance of resonance frequencies and limitations of the mechanism, which are mainly due to the elasticity of the transfer elements (belts). Thus system can operate at higher speed with predictive accuracy.

Trajectories have to be smooth, which means that their first derivative (representing velocity) has to be continuos. Problems occur on the edges when the curve transfers to the next one. In most cases that would cause a step change of first derivative (velocity) introducing vibrations into the mechanism. In rigid mechanisms problems of that kind are solved with the mechanical design, which results in higher material costs [1].

2. G - CODE

G-code is a widespread programming language for industrial tool machines. During the last years many programmers were trained in its use. Programmer can use linear and circular paths together with some additional commands regarding repetitions, zoom, coordinate system placement, etc. Also commands for controlling the laser system are included. Compiler enables the use of absolute and relative coordinates.

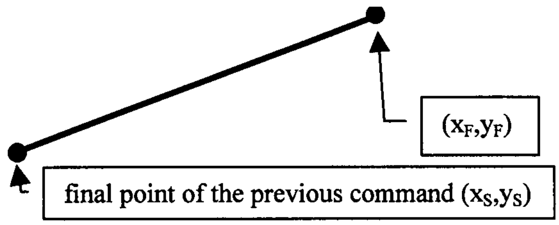


Figure 1: Description of the command G1 - linear path

The main commands used in G-code [2] are:

- G1 Xx_F Yy_F Ff - linear path to (x_F,y_F) with the speed f (Figure 1),
- G2/G3 Xx_F Yy_F Ix_0 Jy_0 Ff - circular move (clockwise / anticlockwise) to (x_F,y_F) with the speed f, circle has the central point (x_0,y_0) (Figure 2),

The final point of the current trajectory is used as the starting point of succeeding one.

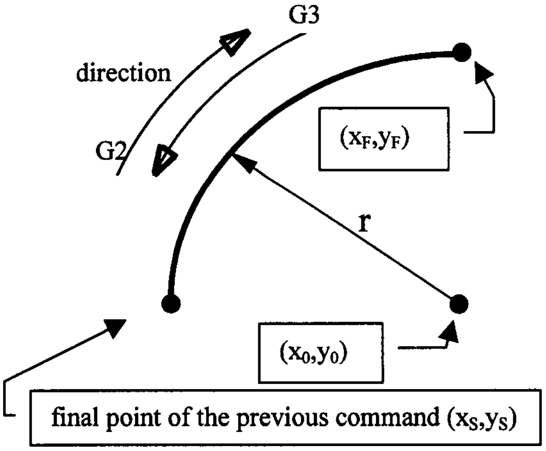


Figure 2: Description of commands G2 and G3 - circular path

3. CHANGING THE SPEED

During the operation speed of movement is changing. In order to assure smooth trajectories, the velocity has to be continuos, which means that it has to change in the continuos manner. In our application algorithm using $\sin^2$ acceleration profile (featured in Figure 3) was used [3,4]. Besides assuring of the smooth movement it is also implementing the avoidance of resonance frequencies and limitations of the mechanism, which are mainly due to the elasticity of the transfer elements (belts). Thus system can operate at higher speed with predictive accuracy.

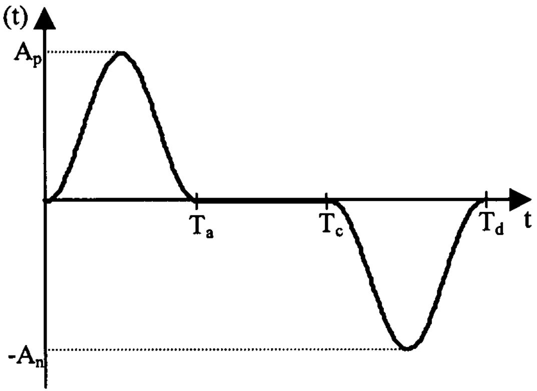


Figure 3: $\sin^2$ acceleration profile

It is very important to mention that the system has no mechanical brakes and all the braking has to be performed electrically.

Two important kinds of limitations are introduced into the planning of speed variations:

- limitations due to the power source and
- limitations due to the resonance frequencies,

which result in the amplitude and time limitations of planned trajectory's accelerations.

In the testing phase the mechanism showed undesired performance when times of acceleration or deceleration phase (T_a and $(T_d - T_c)$ on the Figure 3) were to short. Therefore a lower limit was introduced for those times.

When we talk about speed in G-code program, we can divide programmed trajectories in two major groups:

- trajectories, where the final velocity is different than 0 (machine passes the final point), and
- trajectories, where the final velocity is 0 (machine is stopped in the final point).

In case of passing the final point we only have to consider time and amplitude limitations of the acceleration, whereas when machine is stopped in the final point, also the length of path has to be considered and if necessary, extended. First we shall describe the algorithm used for the trajectories that pass final point.

3.1. Changing the speed when final point is passed

At the programmed speed v_{max} lower than initial speed v_{st} ($v_{max} < v_{st}$), minimal path that can satisfy the time limits is:

$$s_{\min T} = \frac{1}{2} v_{st} T_{a\min} \qquad (1)$$

whereas in case of programmed speed higher than initial one ($v_{max} > v_{st}$) it is:

$$s_{\min T} = v_{st} T_{a\min} \qquad (2)$$

When the programmed path passes that criterion, it has to be tested on the amplitude limitations. The minimal path in this case is:

$$s_{\min A} = \frac{v_{\max}^2 - v_{st}^2}{A_p} \qquad (3)$$

In the case of programmed path shorter than minimal, speed is limited to the value that can be reached with the given maximal amplitude of acceleration:

$$v_{\max} = \sqrt{v_{st}^2 + A_p s} \qquad (4)$$

The obtained speed has to be tested on the time limits again. The minimal path due to the time limitations can be calculated with the following equation:

$$s_{\min T} = \frac{v_{st} + v_{\max}}{2} T_{a\min} \qquad (5)$$

If the test of time limits fails, T_a is set on the T_{amin}. At the programmed speed v_{max} higher than initial speed v_{st} ($v_{max} > v_{st}$), the speed is calculated from the equation:

$$v_{\max} = 2\frac{s}{T_{a\min}} - v_{st} \qquad (6)$$

whereas in other case ($v_{max} < v_{st}$) speed remains unchanged.

When the length of the programmed path satisfies the demands introduced by amplitude limitations, it has to be tested for time limitations for one more time. For the calculation of the minimal path length the equation (5) is used. Changing the speed uses same procedure as in the previous case. First times are calculated and quantized, then v_{max} is balanced with the equation (6).

If all time and amplitude limitations are satisfied, speed can be changed following the program using classical equations for the calculation of time parameters:

$$T_a = 2\frac{v_{\max} - v_{st}}{A_p} \qquad (7)$$

When T_a is smaller than T_{amin} ($T_a < T_{amin}$), it is set on T_{amin} ($T_a = T_{amin}$). Then it is quantized. The next parameter is calculated with the equation:

$$T_d = T_c = \frac{s - \frac{v_{\max} + v_{st}}{2} T_a}{v_{\max}} + T_a \qquad (8)$$

which are again quantized. Balance is provided with the new calculation of T_a, that is quantized again and balanced with the calculation of A_p, which is performed with regard to equation (7).

At this point it is important to repeat that all times are quantized to the sampling time steps after they are calculated.

Figures 4 and 5 describe three typical cases of obtained velocity trajectory.

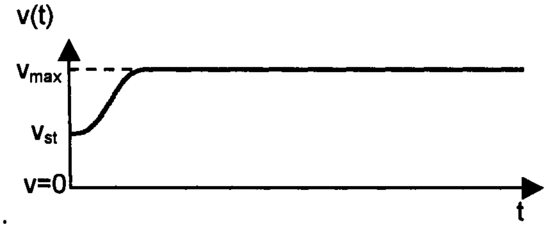


Figure 4: The path trajectory of velocity when $v_{st} < v_{max}$, time and amplitude limitations are not reached (final point is passed)

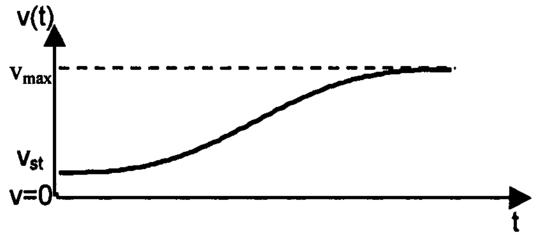


Figure 5: The path trajectory of velocity where time or amplitude limitations are reached (final point is passed)

3.2. Changing the speed at stopping in final point

The algorithm for the second possibility - stopping at the final point has to provide the informations about the acceleration and deceleration phase of programmed trajectory. Thus besides testing demands for the time and amplitude limitations of the acceleration phase, as was the case in former subsection, also tests have to be performed for the deceleration phase, which results in larger algorithm.

The first test is performed considering the minimal path on which movement can be stopped with the given maximal deceleration A_n. The minimal path is calculated with the following equation:

$$s_{\min A} = \frac{v_{st}^2}{A_n} \tag{9}$$

When the programmed path length s is shorter than s_{minA}, it has to be extended to the length of s_{minA} in order to enable the smooth stopping of the mechanism. This new value has to be further tested on the time limitations.

The minimal path allowed by the time limitations s_{minT} is given following the equation (1). If the trajectory obtained with the equation (9) is shorter, it has to be extended to this value. Since in this case there is no acceleration phase ($T_d = 0$) and also no phase of constant velocity ($T_c = 0$), only deceleration phase remains. Its duration is calculated with:

$$T_d = 2\frac{v_{st}}{A_n} \tag{10}$$

T_d is quantized to the sampling time and balance is performed through the new acceleration with regard to equation (10).

At the satisfied amplitude limitations the trajectory again has to be tested on time limitations for which the equation (1) is used. If time limitations are reached, T_d is set to T_{amin} ($T_d = T_{amin}$), quantized and after that path is extended to minimal possible, which was calculated with the equation (1). Again there is no acceleration phase ($T_a = 0$) and also no phase of constant velocity ($T_c = 0$).

If the programmed trajectory has successfully passed all previous tests to the various limitations, the existence of acceleration and deceleration phase is assured.

First the duration of deceleration phase and phase of constant velocity ($T_{d-a} = T_d - T_a$) is calculated.

$$T_{d-a} = 2\frac{v_{\max}}{A_n} \tag{11}$$

If it is lower than T_{amin}, it is extended to T_{amin} ($T_{d-a} = T_{amin}$) and quantized, which is balanced with the following equation:

$$A_n = 2\frac{v_{\max} - v_{st}}{T_{d-a}} \tag{12}$$

After that duration of acceleration phase is calculated with the equation (7). In the case when it is lower than T_{amin}, it is extended to T_{amin} ($T_a = T_{amin}$) and quantized.

The obtained value of T_a is tested on the time limitations for one more time. It has to be lower or equal to minimal value T_{aminv} in order to assure the reaching of velocity v_{max}.

$$T_{a\min v} = 2\frac{s + \frac{v_{st}^2}{A_n}}{v_{st} + v_{\max}} \tag{13}$$

If T_a is higher than T_{aminv} ($T_a > T_{aminv}$), solution is different for different relation of initial and programmed velocity.

For programmed velocity lower than initial one the following procedure is used.

First deceleration time is calculated:

$$T_d = \frac{s + \frac{v_{st}^2}{A_{n\max}}}{v_{st}} \tag{14}$$

where A_{nmax} is maximal possible value of deceleration amplitude. T_d is quantized to the sampling time and balanced with the deceleration amplitude (with respect to equation (14)).

The time parameter for the phase of constant velocity is calculated as follows:

$$T_c = T_d - 2\frac{v_{st}}{A_n} \tag{15}$$

There is no acceleration phase ($T_a = 0$). T_c is used for the balance due to quantization.

For programmed velocity higher or equal than initial one ($v_{max} \geq v_{st}$) we use the following procedure.

First the programmed velocity is limited with the following equation:

$$v_{\max} = \frac{2s - v_{st}T_a}{T_a + T_{d-a}} \tag{16}$$

The balance for quantized times and new velocity is performed in next step through the adaption of acceleration amplitudes. Equation (8) is used for the calculation of A_p and following equation is used for the calculation of A_n:

$$A_n = 2\frac{v_{\max}}{T_{d-a}} \tag{17}$$

There is no phase of movement with the constant velocity ($T_c = T_a$).

The last possibility is when all limitations can be reached. In this case time parameter for the deceleration phase is calculated as:

$$T_d = T_{d-a} + \left(\frac{s - \frac{v_{st} - v_{\max}}{2}T_a - \frac{v_{\max}^2}{A_n}}{v_{\max}} \right) \tag{18}$$

The obtained value is quantized on the sampling time and the time parameter for the constant velocity phase can be calculated.

Figures 6 to 8 describe three typical cases of obtained velocity trajectory.

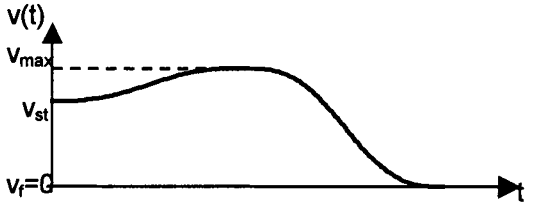


Figure 6: The path trajectory of velocity for limited v_{max} when $v_{st} < v_{max}$, time or amplitude limitations are reached (stopping in the final point)

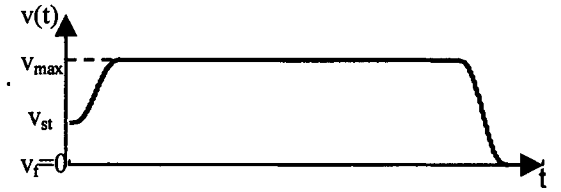


Figure 7: The path trajectory of velocity when $v_{st} < v_{max}$ and v_{max} is not limited, time and amplitude limitations are not reached (stopping in the final point)

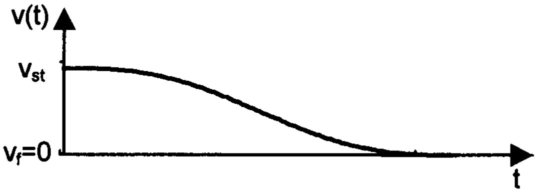


Figure 8: The path trajectory of velocity where time or amplitude limitations are reached and $v_{st} > v_{max}$ (stopping in the final point)

4. INSERTING CIRCULAR SPLINES

In non-rigid systems problems originating from momentary changes of movement direction can be partially solved with the introduction of additional curves, that is if a contour transfers to another one with a sharp corner, circular arc is inserted as a connection enabling a smooth movement. In rigid mechanisms problems of that kind are solved with the mechanical design, which results in higher material costs.

The inserted circular arc starts on the first and ends on the second programmed contour. At the end points it has the first derivative equal to the one of the programmed contour. Thus a continuos velocity is assured. The radius of inserted circular arc can be chosen by the programmer.

Results obtained without the insertion of circular splines are featured in Figures 9 and 10. Discontinuities occur in the points where contours contact. Only a smaller particle, lying on the point of transfer from linear to circular contour, was chosen, since the effects of interest can not be shown with the complete path.

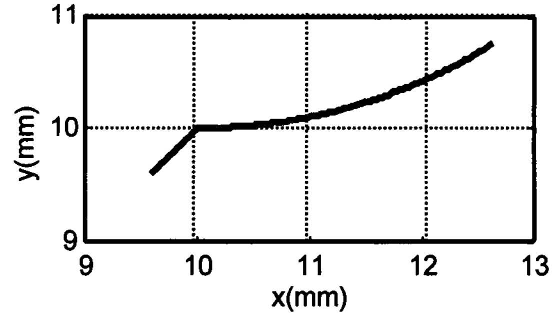


Figure 9: Detail of the programmed path

As featured on the Figure 9, contours touch with the angle different than 0, which means that direction of movement is changed in non-continuos manner.

Figure 10 shows velocity trajectories. It is important to stress that velocity trajectories are not continuos.

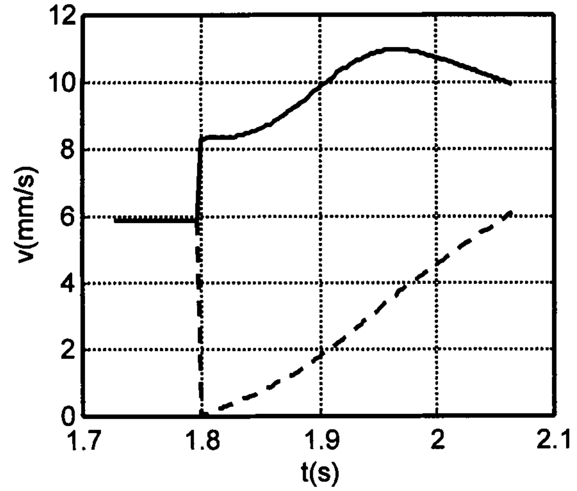


Figure 10: Reference velocity in x-axis (solid line) and y-axis (dashed line) for the path described in Figure 9

An example for the introduction of circular arcs into the programmed path is described in the Figures 11 and 12.

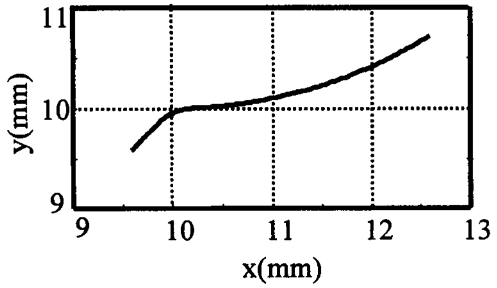


Figure 11: Path in Figure 9 with inserted spline

As it is shown in the Figure 12, velocity trajectory is continuos, which is not the case for the velocity trajectory without inserted circular splines (Figure 10). However, this procedure is not successful for the smoothing of acceleration trajectories.

5. CONCLUSION

Since it is well known from physics that no mass (object) can change its speed or direction in the infinitive short time, the actual trajectory could not match the programmed trajectory completely. With the use of inserted circular splines programmer obtains the possibility of determining deviations in such cases. Since acceleration is not continuos, an additional trajectory filter has to be used in order to ensure the precise trajectory tracking. Thus also additional protection from the occurrence of resonance frequencies can be obtained.

Algorithms for the assurance of the smooth movement of the non-rigid system were used on the real system.

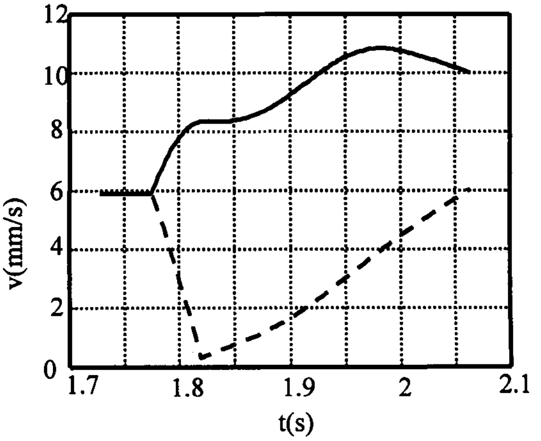


Figure 12: Reference velocity in x-axis (solid line) and y-axis (dashed line) for the path described in Figure 11

REFERENCES

1. K. S. Fu, R. C. Gonzalez, and C. S. G. Lee, Robotics - Control, Sensing, Vision and Intelligence McGraw-Hill International Editions, 1988.
2. *MAHO-Universal-Fräs- und Bohrmaschine mit der Steuerung CNC 432*, Schulungsunterlagen, 1988
3. I.Van Aken and H.Van Brussel, On-line robot trajectory control in joint coordinates by means of imposed acceleration profiles, Robotica, June 1988, pp. 185-195
4. M. Terbuc, K. Jezernik and G. Haßler, Cartesian Space Path Tracking with Transputer Robot Controller, 25th International Symposium on Industrial Robots, Hannover 1994, Proceedings pp. 411-418.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Adaptive Control of Mechanical Processes: Brakeforming of Metal Sheet as an Example

L.J. De Vin[a] and U.P. Singh[b]

[a]Department of Engineering Science, University of Skövde,
Box 408, 541 28 Skövde, Sweden, e-mail leo@ite.his.se

[b]Advanced Forming Technology Group, University of Ulster at Jordanstown,
Newtownabbey, Co. Antrim BT37 0QB, UK

In air bending, the variations in mechanical sheet properties present problems as regards the calculation of the required punch displacement. As a result, bend angles are often not accurate enough to meet today's high tolerance requirements. This problem can be reduced with the use of adaptive control methods based on in-process measurement of bending characteristics. Adaptive control methods based on the measurement of punch displacement and bend angle and/or the process force are discussed. An improved method for the interpretation of process force - punch displacement data is presented. This method incorporates the use of an analytical bending model which is used to predict the sheet behaviour in off-line simulations. Therefore, a better data interpretation is achieved. The method also reduces the data processing time during the bending cycle. In this way, it is possible to eliminate some of the disadvantages of other methods.

1 INTRODUCTION

In mechanical manufacturing processes, there can be problems due to variations in the behaviour of the equipment used, the process itself or the material being processed. Process monitoring can be applied as a safety measure. For instance, in machining, the cutting force can be measured and compared with a predicted value. If the two values differ too much, the tool may be broken or there may be a collision. Adaptive control aims at optimising a process or at generating output between required predefined limits. Ideally, the output parameter that must be kept under control is measured, and the result is used to determine required input for the process. However, adaptive control of brakeforming is more complex.

2 BRAKEFORMING

In brakeforming, a sheet of metal is clamped between a punch and a die. Subsequently, the punch is lowered into the die and the a bend is made. When the punch is retracted, the sheet shows spring-back behaviour due to the material's elasticity. The main processes applied are closed-die bending and air bending (also called free bending, Figure 1). In air bending, the punch is lowered to a given, usually precalculated, position. In this way, it is possible to bend a sheet to different angles or to accommodate different spring-back behaviour during unloading. Air bending thus offers the potential flexibility of producing different products (with respect to geometry, sheet thickness and material) with the use of only a limited number of tool sets, thus reducing the need for tool changes. Usually, the sheet radius in air bending is larger than the punch radius. When they become equal, this is called “wrap-around” and the corresponding punch penetration is called “wrap-around point”. One of the major problems in air bending is that the material behaviour may show considerable variations which makes it difficult to obtain a high accuracy in bend angle. Adaptive control of the process aims at solving this problem.

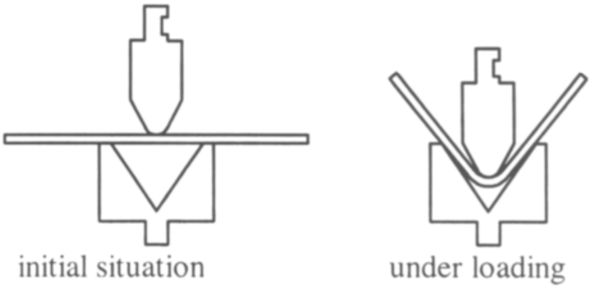


Figure 1: Air bending

3 PRINCIPLE SOLUTIONS FOR ADAPTIVE CONTROL IN BRAKEFORMING

3.1 Measurement of the bend angle

The most elementary approach would be, to measure the bend angle continuously and stop the punch displacement when a given value is reached. Since continuous measurement of the bend angle can prove to be difficult, an alternative would be to proceed to a certain safe position, measure the bend angle, and then calculate the required remaining punch travel with the use of an analytical process model.

However, these methods share one common problem, namely that variations in spring-back angle can not be compensated for. These variations in spring-back angle can be significant and sometimes difficult to predict. This is due to the change of Young's modulus of elasticity under deformation, variations in stress conditions for different products (eg strip or wide sheet), and variations in local geometry, in particular smallest bend radius.

The process force can be measured in an attempt to overcome the spring-back problem. Towards the end of the bending cycle, the punch is retracted until the load is a given percentage of the initial load and a partial spring-back is measured. One problem here is, that the unloading - spring-back behaviour is non-linear [2].

3.2 Measurement of the process force

A more fundamental solution lies in the measurement of the process force and the generation of a force-displacement curve. This curve can subsequently be analysed and from this, the moment-curvature relationship, valid under the deformation conditions at hand, can be derived. This relationship can be used in a bending model, allowing the determination of the sheet geometry.

4 SENSORY EQUIPMENT

4.1 Measuring the bend angle

The bend angle can be measured with the use of mechanical devices or optical equipment (Figure 2). Mechanical devices can be a probe, measuring the distance between the sheet and a fixed point on the press brake, or a faceted rotary disk. This mechanical equipment is relatively simple, but some reported problems include: wear of the equipment and the tools; increased tool changeover times, and interference with the bending process. In particular, problems with set-up selection can occur, depending on the geometry of the sheet metal parts. The use of a probe has the additional disadvantages of having a limited range and problems with holes in the sheet.

Optical methods include the use of overhead camera's mounted in or near the punch and retractable sensors mounted near the die. Overhead camera's have the disadvantage that parts of the sheet may obscure the part of which the angle has to be measured. Lutters [1] describes the use of a small retractable sensor near the die. This device has an emitter and two sensors, and rotates over a small angle. The determination of the bend angle is based on two reflectivity peaks. As a result, surface quality becomes less important and even perforated sheet gives fair results. However, when bend legs are short, as is common for many components such as panels and sashes, the sensor can not be used.

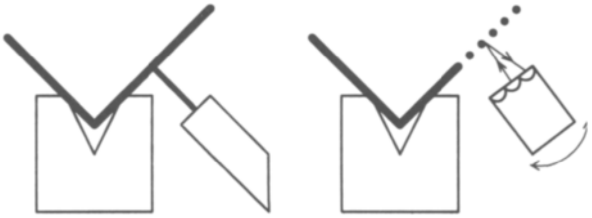


Figure 2: Mechanical probe & optical sensor

4.2 Measuring the process force

One way to measure the process force is to glue strain gauges on a tool (usually the die). However, this method is not well suited for industrial use. During tool changes, the strain gauges must be disconnected and the strain gauges on the new tool must be connected. Another problem is that the characteristics of gauges may differ and therefore, each set must be calibrated and its data stored in the adaptive control software.

Lutters [1] describes the use of a dowel type piezoelectric sensor that is inserted into a hole in the die. However, the signal obtained from the sensor depends on the horizontal and vertical position of the sensor in the die. Even for one size of die, finding the best position seems to be a trial-and-error procedure. In order to avoid problems with calibration and also for economical reasons, only one sensor is used which may be inserted in different dies. This requires a consistent good quality fit and problems with wear and dirt may be expected.

A third way of measuring the process force is with the use of a pressure transducer in the hydraulic system. The major advantage is that this type of force measurement does not hamper the operation of the press brake in any way and that tool changes do not present problems. A disadvantage is the limited range of pressure transducers and the wide range of forces.

However, Kamlage [7] has demonstrated that after careful calibration, two transducers suffice to cover a range of forces corresponding with sheets ranging from narrow strip to full width of the press brake. Important advantages of this type of sensor is that they are robust, have no trailing leads near the tools, and that they are affordable. The latter is important since the budget for additional equipment is usually limited to 5% to 10% of the total press brake costs.

5 METHODS BASED ON ANGLE-DISPLACEMENT MEASUREMENT

Optical measurement with an overhead camera is used by Heckel [3]. When parts of the sheet metal product cover the bent face, the bend angle can not be measured. This is not a problem when the mechanical properties within one sheet do not vary too much. In that case, the punch displacement calculated for previous bends can be used. This is only possible when bends with the same bend angle and orientation with respect to the rolling direction have been bent previously. The method proposed by Heckel is based on an incremental bending procedure. Each step consists of a loading phase and an unloading phase during which the punch is retracted almost to the release point. When the resulting bend angle is still too small, a new bending cycle with a larger punch penetration is executed. This procedure is repeated until a satisfactory bend angle is obtained.

The accuracy that can be achieved with the use of this method will depend on the die opening and the minimum step for the punch displacement. For a die opening resulting in a sensitivity of 15°/mm and a smallest increment of 0.01mm for the punch displacement, Heckel claims an accuracy of the bend angle of ±0.1°.

Lutters [1] describes a method to measure the bend angle optically with the use of a retractable sensor mounted near the die. He uses modular tools, with a piezoelectric force-sensor in a small tool module. This method seems to give good results, even for coated, dirty or perforated sheet, but problems with springback prediction are also reported. Especially materials with a large Young's modulus and thus a small springback angle are reported to give problems. This may be due to insufficient sensitivity of the force sensor or problems with the fit between the sensor and the hole in the die. The non-linear behaviour of the springback presents additional problems.

6 METHODS BASED ON FORCE-DISPLACEMENT MEASUREMENT

The process force can be measured in such a way that the measuring equipment does not influence the normal operation of the press-brake, especially when pressure transducers are used for this purpose. The major problem in the force-displacement method is the necessity to transform the measured data into moment-curvature or stress-strain relationships. The generation of the moment-curvature relationship is sufficient for adaptive control purposes. The generation of stress-strain relationships can be useful to obtain material characteristics under bending conditions that can be used for off-line simulation of bending and related processes. Stelson [4,5] has investigated the interpretation of force-displacement curves whereas Pickering [6] and Kamlage [7] have carried out research on in-process data logging.

The principle of generating a moment-curvature relationship from the force-displacement data is explained in Stelson [4]. The bending moment can be calculated from the process force and the effective die opening. For larger bend angles, the effect of friction and the contribution of the horizontal component of the force have to be taken into account. In addition, the effective die opening decreases during bending due to the changing contact between the sheet and the die. For the prediction of these effects, Stelson [4] assumes that the centre line of the sheet is a straight line from the punch to the die. For the calculation of the bending moment, this results in Eq (1) where F is half the process force. The length L is half the die opening. The calculation of the curvature requires differentiation as shown in Eq (2).

$$M = FL \tag{1}$$

$$K(M) = \frac{1}{L^2}\frac{1}{F}\frac{d}{dF}\left[F^2 Y(F)\right] \tag{2}$$

Stelson indicates that for the larger bend angles, L must be replaced by half the effective die opening L' and F must be replaced by F'. F' is an equivalent force which, applied with bending arm L' results in the bending moment under the punch. The force F' is larger than F or F_V due to the contribution to the bending moment of the horizontal component F_H of the force between the die and the sheet (see Figure 3).

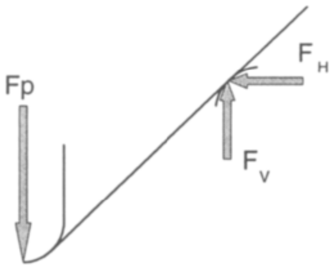


Figure 3: Forces acting on the sheet

With the use of this method, Pickering [6] obtained bend angles within a band width of 0.4°. Even though the average punch displacement in his experiments was larger than the pre-calculated value, the average bend angle was smaller than the objective angle of 90°. Pickering does not give an explanation for this but it is probably due to the change of Young's modulus under deformation, resulting in a larger spring-back angle. Its effect seems to have been smaller than usual which could indicate a significant hold time of the punch displacement, allowing the sheet to settle.

Kamlage [7] has carried out work on the instrumentation for in-process measurements of the process force and the punch displacement. When the measured data are to be used without curve-fitting, the incoming signals must be filtered to remove any excess noise. A problem is that the beginning of the curve (in the elastic zone and near the elastic - elasto-plastic transition) changes when high order filtering is used. This problem can be avoided by starting the data logging after the transition point. This transition point can be calculated with a sufficiently high accuracy.

7 AN IMPROVED METHOD FOR THE INTERPRETATION OF FORCE-DISPLACEMENT DATA

The method developed by Stelson has two drawbacks: (i) The method requires extensive data processing during the bending cycle and (ii) The model used to predict the actual die-opening, the actual punch penetration and the direction of the forces is very simple. This section describes how these two problems can be solved with the use of an analytical bending model.

The bending model described by De Vin [8] is a plane strain bending model, based on the determination of the moment-curvature relationship. Using this relationship, the bend angle under loading conditions can be calculated. With the use of elastic beam theory, the bend angle after spring-back can be calculated. This bending model can be used to simulate the bending process off-line. Nominal values for the mechanical properties of the sheet material as obtained from a tensile test or given by the material supplier are used during this simulation. Due to the more accurate description of the sheet geometry as compared to the model used by Stelson, a better prediction of the effective die opening and the direction of the forces is obtained. Furthermore, it is possible to generate a file with multiplication factors off-line. With the use of these factors, the major constituent parts of Eqs (1) and (2) can be calculated with a simple multiplication. The curvature can subsequently be calculated from Eq (2) using numerical differentiation, with is also a very fast operation. In this way, the required computing times during bending can be reduced to a minimum.

Figure 4 illustrates that the effective punch penetration Y'' and half the effective die opening L'' differ from the corresponding dimensions Y' and L' as estimated with the circular-straight approximation. Furthermore, the directions of the normal and tangential forces differ from the directions obtained from the circular-straight approximation. As a result, the interpretation of the in-process measured data is not correct. Off-line simulations using the bending model give a more accurate interpretation.

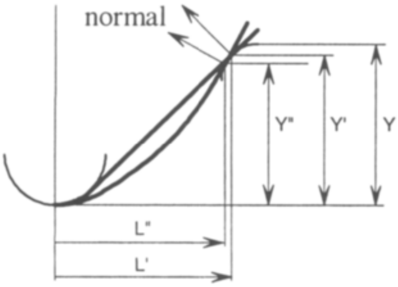


Figure 4: Nominal, circular-straight and true dimensions

Figure 5 shows the multiplication factors generated for a 2mm mild steel sheet and a ModufixÔ OX-5 die. The non-linear behaviour of *Pfac* and *Mfac* is clear, especially for the larger punch displacements where the difference between the actual sheet geometry and a circular-straight approximation becomes larger. The difference between the actual bending moment and the moment calculated from the circular-straight approximation can become as large as 20%. Together with errors in Y' and L', this results in problems regarding the prediction of the wrap-around point and calculation of the sheet curvature in general.

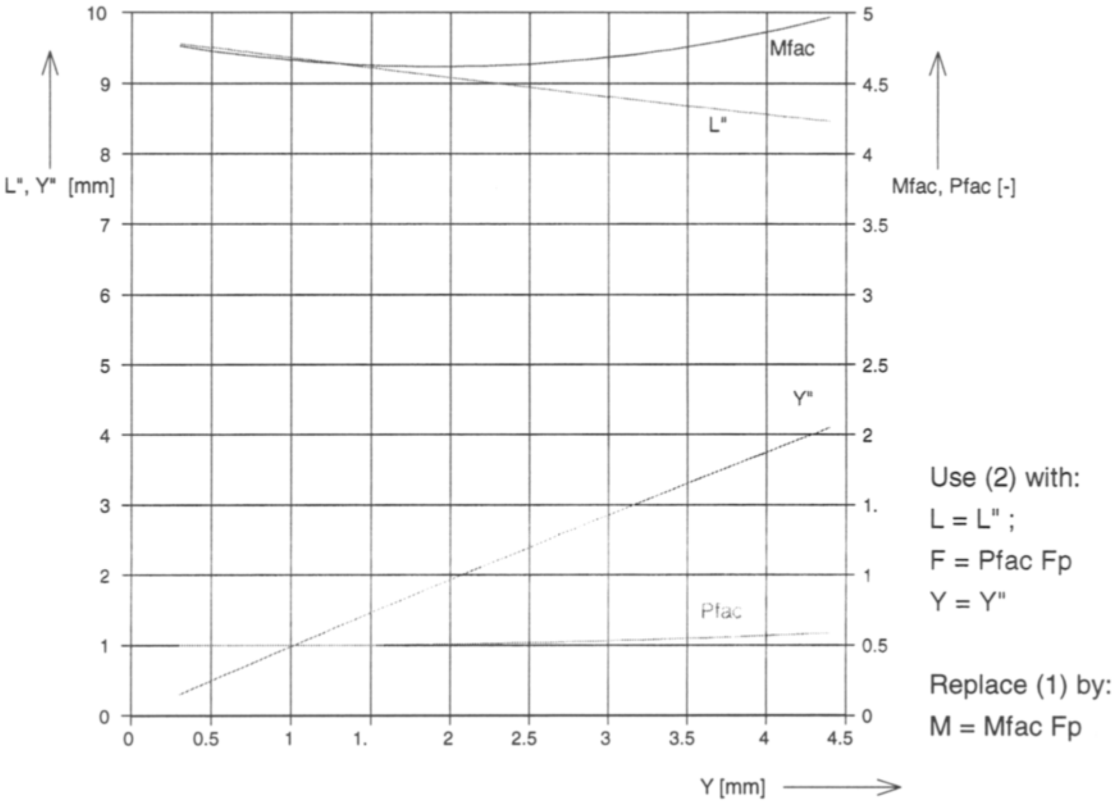


Figure 5: Graphic representation of the multiplication factors

Both Stelson and Pickering advocate curve fitting of the force-displacement curve, which results in a mathematical function for the moment-curvature relationship which permits extrapolation to other curvatures than those calculated. However, such extrapolations have a very limited validity and use. With the off-line simulation discussed above, it is possible to generate multiplication factors for the whole punch stroke and the data can be processed until a wrap-around curvature is obtained. This also solves the problem that measurements taken after the wrap-around point may inadvertently affect the accuracy of the results, which can be the case when curve-fitting is used to obtain the moment-curvature relationship.

8 SPRING-BACK

The spring-back behaviour which causes problems in other methods, is less of a problem when the force-displacement method is used. The method provides information to calculate not only the overall sheet geometry but also the local curvatures of the sheet which allows a better prediction of the spring-back. In particular, the non-linear behaviour can be explained and calculated with the use of the analytical model. The only problem that remains is the change in Young's modulus, which can be estimated to be about 20% for mild steel [8]. A further improvement may be to perform a punch retraction cycle which can provide data to estimate the change of Young's modulus even better.

9 THE GENERATION OF STRESS-STRAIN CURVES

It is also possible to obtain the stress-strain relationship from the moment-curvature relationship. The tangential engineering strain can be calculated from the sheet curvature. The tangential stress can be found by a second differentiation (Nadai [9]). However, the moment-curvature relationship provides sufficient input for the bending model and adaptive control of the process. The stress-strain relationship can be used for the analysis of bending and bending related processes, for instance rubber pad forming of flanges with slightly curved heel lines.

10 CONCLUSIONS

The material behaviour in air bending is a relatively unpredictable factor and therefore, the accuracy of the process can be improved significantly by using adaptive control. However, some methods may hamper the operation of the press brake or suffer from limited applicability. Methods based on the measurement of force-displacement data do not have this drawback. However, extensive data processing may be required and the interpretation of the data can be inaccurate. A method to overcome these problems with the use of an analytical bending model has been presented. Instead of being a mere automated trial-and-error method, this method is based on understanding of the process behaviour and offers a more fundamental solution to the problem.

REFERENCES

1. Lutters, D., Streppel, A.H., Kroeze, B. & Kals, H.J.J.,*Adaptive press brake control in air bending*, Proc. of the Shemet97 Conference, Belfast, pp 471-480, 1997.
2. Finckenstein, E. von, Rothstein, R., *Gesenkbiegen von Blechen. Automatische Rückfederungskorrektur*, Industrieanz., Vol. 24, pp 19-20, 1988.
3. Heckel, W. & Franke, V., *In-process Measurement of Bending Angles*, Proc. of the SheMet95 Conference, Birmingham , pp 272-281, 1995
4. Stelson, K.A., *Real Time Identification of Workpiece-Material Characteristics from Measurements during Brakeforming*. J. of Eng. for Ind. - Trans. of the ASME, Vol. 105, pp 45-53, 1983.
5. Stelson, K.A., *An Adaptive Pressbrake Control for Strain Hardening Materials*, J. of Eng. for Ind. - Trans. of the ASME, Vol. 108, pp127-132, 1986.
6. Pickering, S., *Intelligent Processing of Materials*, Journ. of Mater. Proc. Technol., 36, pp 447-465, 1993
7. Kamlage, G., *In-Process-Characterisation of Material Behaviour in Free Bending*, Student Thesis, University of Ulster at Jordanstown, 1994
8. De Vin, L.J., Streppel, A.H., Lutters, D. & Kals, H.J.J., *A Process Model for Air Bending in CAPP Applications*, Proc. of the Shemet94 Conference, Belfast, pp 17-28, 1994.
9. Nadai, A., *Theory of Flow and Fracture of Solids*, 2nd ed, McGraw-Hill, New York, Vol 1, pp 356-359, 1950

NOTATION

Fp	process force
F	bending force (idealised)
F'	equivalent force (circular-straight approximation)
F"	true equivalent force
F_V, F_H	vertical and horizontal contact forces
L	half die opening (idealised)
L'	effective half die opening (circular-straight approximation)
L"	true effective half die opening
K	sheet curvature under the punch
M	bending moment under the punch
Mfac	multiplication factor to calculate the bending moment
Pfac	multiplication factor to calculate the equivalent force F"
Y	punch displacement
Y'	effective punch penetration (circular-straight approximation)
Y"	true effective punch penetration

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A piezoelectrically driven wire feeding system for high performance wedge-wedge-bonding machines

A. Henke, M. A. Kümmel, J. Wallaschek

Heinz Nixdorf Institut, Universität-GH Paderborn, Fürstenallee 11, 33102 Paderborn, Germany

In this paper a new wire feeding system for an aluminium wire bonding machine and its development is presented. The system is driven by a piezoelectric actuator, which replaces the electromagnetic actuators used in present bonding machines. To minimise the time used to develop the new system, several models of the system have been derived and simulations have been made, to optimise the systems performance. After the wire feeding system has been manufactured, experiments have been made to validate the results conceived from the simulations. The performance improvement to the currently used feeding system is discussed.

1. INTRODUCTION

Only ten years ago the electronic part of a machine still took up a significant amount of its volume. The miniaturisation of electronic parts provided the chance to reduce a machines dimensions but also to equip a machine with more 'intelligence' up to decentralised information processing, which is located there where it is actually needed.

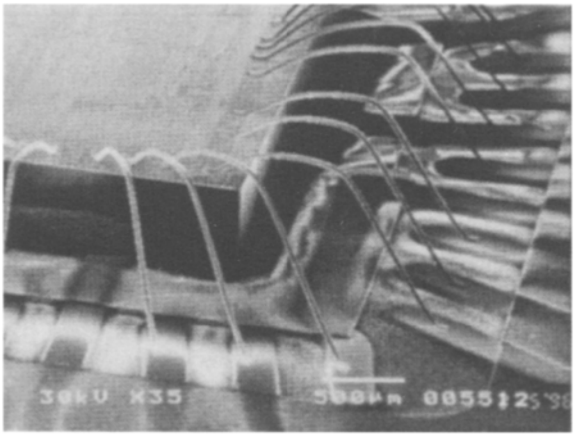


Figure 1. Integrated circuit connected to the board by COB-technique.

The miniaturisation in the field of microelectronics imposes new demands on production processes especially in posterior processing. One of the most common production techniques used to apply integrated circuits on a circuit board is the Chip-On-Board technique (COB). Here the integrated circuits are attached directly onto the circuit board. The formerly used sockets and housings are spared (fig. 1).

Modern applications in microelectronics require high accurate positioning and high speed bonding machines. The process of bonding is one of the last production steps in the manufacturing of microelectronic modules. Therefore it is important to ensure quality and reliability of the process to avoid the loss of the value arisen from the more expensive processes applied before. One step to fulfil these requirements was the new development of the wire feeding and clamping system of a bonding machine.

2. ALUMINIUM WIRE BONDING

The electrical connections between chip and board are generated by bonding machines. The process of bonding essentially bases on an ultrasonic friction welding process. Very thin wires of diameters 17.5 μm to 80 μm are used for the connection and welded to chip and board at both ends (see also[1]).

2.1. Welding Unit

The welding unit used in a bonding machine is placed in the bond head, which is positioned by a three dimensional Cartesian robot with an additional rotary axis. The main parts of the welding unit are the ultrasonic vibration transducer and the wedge, which is the actual welding tool. The wedge is attached to the tip of the transducer (fig. 2). A piezoelectric actuator excites the transducer to longitudinal vibration, which causes transversal vibration of the wedge in the range of 100 kHz. The resulting relative displacement between the tip of the wedge and the circuit is used to weld the wire onto the circuit board.

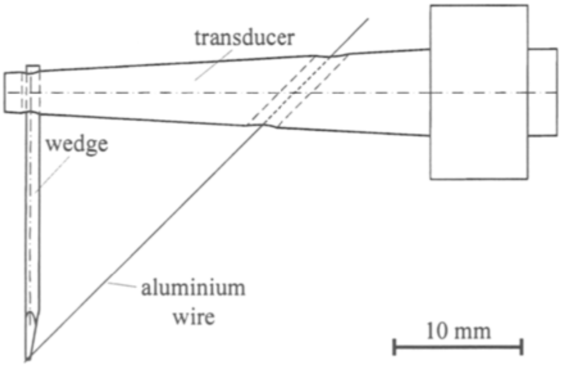


Figure 2. Welding unit of an aluminium wire bonding machine.

2.2. Bond Cycle

During a bond cycle two points (source and destination) are connected with the aluminium bonding wire (see fig. 3).

For the first bond the welding unit is positioned above the first welding point (source). After the touchdown of the wedge on the substrate, the welding process for the first bond starts. Thereafter the welding unit is driven in the direction to the second bond, while the wire clamp remains open so that the wire is able to slide through it. After a designated distance the wire clamp fixes the wire and the welding unit is driven a certain trajectory to form the loop, which is the arched wire connection of the two bonds. After the touchdown and the welding process at the destination point the wire has to be torn off close to the welded zone. This is realised by clamping the wire and moving the clamping system backwards. At this point of the bond cycle there is no wire left below the bond flat of the wedge. To set the welding unit up for the next bond cycle the wire has to be fed underneath the wedge [2]. These two movements of the clamping system, feeding the wire underneath the wedge and tearing it off after the connection is done, are the strokes the wire feeding system has to perform. In figure 3 the bonding process is shown during the welding phase at the second bond.

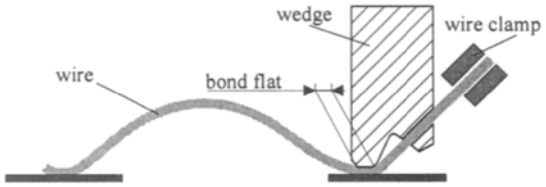


Figure 3. Bond connection

2.3. Problems in present bonding machines

In present day bonding machines the wire feeding and clamping system is driven by electromagnetic actuators. Their advantage is their large displacement. However, to become fast, accurate, and robust they work against mechanical stops. As a consequence, the system is disturbed by an impact during every single positioning process, which results in interfering vibrations. These vibrations decelerate the bond cycle, because for technological reasons the machine has to wait until the vibrations have vanished.

Another disadvantage resulting from the use of mechanical stops is the awkward mechanical adjustment process, which is required after rearranging the system for a new wire diameter or different stroke length. This negatively effects the handling convenience of the bonding machine. The precise adjustment of the feeding system (the mechanical stops) in the range of micro meters is nearly impossible.

Another problem results from the distinct wear of the electromagnetic actuators. This forces the operator to frequently adjust and clean the actual feeding system and to replace it soon.

To take future requirements into account, attention has to be paid to the following. In modern electronic modules the microelectronic components are packed very closely. This requires a small welding tool for the bonding machine to be able to generate fine and accurate connections. The size of the whole welding unit, however, depends on the size of the welding tool. Therefore the size of the whole welding unit has to be reduced. The reduction of the size of the welding unit is also necessary to achieve higher bonding frequencies for an acceleration of the bonding process. Hence, there is less space for the wire feeding and clamping system. Presently electromagnetic actuators are not available with same performance and smaller size. Therefore a new concept to drive the wire feeding and clamping system has to be developed.

3. PARALLELOGRAM ACTUATOR

The described problems can be solved by using a piezoelectric actuator. Because of its energy density a piezoelectric actuator is predestined for the tasks described above. Besides it can be driven in an open loop control with still being very accurate in positioning, because of its high stiffness. Another advantage of piezoelectric actuators is their high output force. But on the other hand piezoelectric actuators have one significant disadvantage, their small output displacement. Therefore an appropriate and system adapted displacement amplification had to be designed.

3.1. Transmission Design

The requirements for this transmission are a high transmission ratio, a backlash free motion, and a compact and light design. After comparing several design concepts like lever transmission, general four-bar linkage, and others a parallelogram transmission was found to be the most appropriate for this application. It has a very compact design and is able to provide high transmission ratios. In addition to that it provides a quasi linear output displacement. The actuator, which drives the parallelogram frame, acts like a diagonal with variable length (see fig. 4a).

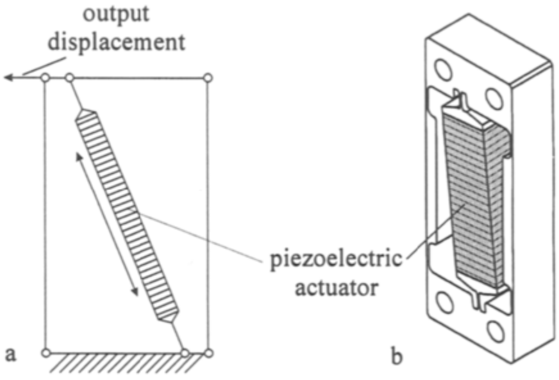


Figure 4. a: parallelogram transmission for displacement enlargement, b: parallelogram actuator CAD model.

Transmission rations of 1:10 and higher can easily be achieved with this concept. For the application of a wire feeding system of a bonding machine a displacement of approximately 300 µm is needed. To keep the transmission ratio low and the size of the piezoelectric actuator small at the same time, the maximum transmission ratio was set to 1:10. Hence, to reach a 300 µm output displacement, a piezoelectric actuator with a displacement of 30 µm is needed. The size of the chosen multilayer actuator is 26x8x8 mm^3 with an output displacement of 30 µm. A CAD model of the parallelogram actuator (parallelogram frame plus piezoelectric actuator) is shown in fig. 4b.

3.2. Geometric Optimisation

A finite element analysis showed, that the flexure hinges connecting the piezoelectric actuator with the parallelogram frame have to sustain the highest stress. Therefore they have to be designed very properly. The load of several contours for the hinges have been calculated with FE analysis. The different hinge shapes were circular, filet type and elliptic (see also [3]). Because of the high stress resulting from the bending load in the flexure hinges, they have been designed in a narrow elliptic contour, that is elastic in bending direction. The final shape of the hinges is shown in figure 5.

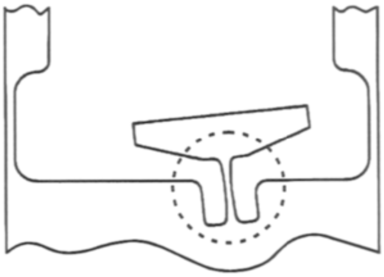


Figure 5. Flexure hinge with elliptical contour.

Because of their slender shape the hinges have to be checked on buckling. In fact there are two load cases which can cause buckling. The first one occurs if the piezoelectric actuator is driven to extend itself and at the same time the parallelogram frame is falsely restricted in its motion. This causes a pressure load in the hinges. The second load case might occur during attaching the feeding system to the bonding machine. Pushing the system in negative displacement direction again causes a pressure load in the hinges.

The system has been optimised in terms of maximum stress (flexure hinges) and buckling. The design safety factors are 3 and 6 respectively.

3.3. Dynamic Investigations

To check the dynamic performance of the system, several dynamic simulations have been made. The requirements for the system are not only to get a fast and accurate system, but also to drive it in open loop control only.

In a first step the mechanical subsystem of the parallelogram actuator can be approximated as a single degree of freedom (SDOF) system, which is represented be:

$$m\ddot{x} + d\dot{x} + cx = f(t) \tag{1}$$

Where x is the displacement of the system, m the effective mass, d the damping coefficient, and c the systems stiffness. The required parameters were obtained from FE analyses. To get the systems stiffness, a force has been applied to the system an the displacement has been calculated. The damping coefficient was estimated based on experiences from similar systems. If the two coefficients c and d are known, the last coefficient (m), needed to describe the system, can be calculated from the systems natural frequency. This was obtained from a dynamic FE analysis (fig. 6).

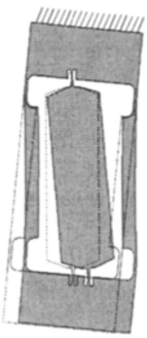


Figure 6. First mode shape at 684 Hz.

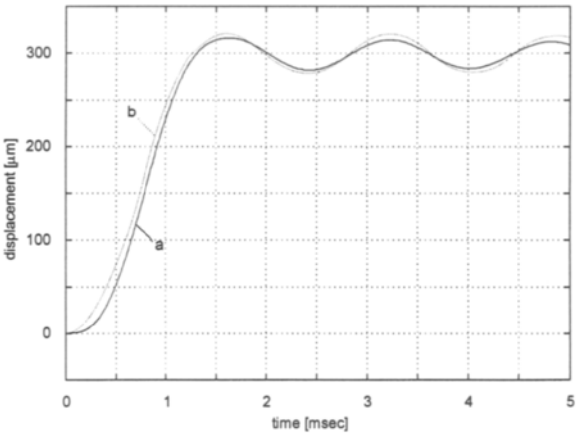


Figure 7. Response of a ramp input: a: System represented as single degree of freedom system; b: FEM simulation.

To check how exactly the SDOF model represents the systems behaviour the response of a ramp input to the SDOF model (fig. 7a) was compared to the FE simulation (fig. 7b). The comparison of these two responses show, that the basic dynamic behaviour of the system is approximated well by the SDOF model.

3.4. Control Strategies

To find an appropriate open loop control to drive the system, several input strategies have been tested. These strategies are as follows:

- Double step:
  The system is excited by two steps. The first one causes an overshooting to the required position. Exactly to the time when the maximum position is reached, the second step keeps the system, which is currently resting, in its position.
- Ramp:
  A ramp, which is adapted to the systems first natural vibration period, is used as input.
- Sine:
  The input of the system is a half period of a sine wave, which is also adapted to the systems first natural vibration period.

The according graphs of the input voltage of the presented control strategies are shown in figure 8a. The response of the system on these input strategies strongly differs in the time needed to reach the required position. For an ideal SDOF system with no damping these times are 0.5 T, T, and 1.5 T respectively to the order the strategies are mentioned above. Where *T* is the vibration period of the systems first natural frequency (see fig. 8b).

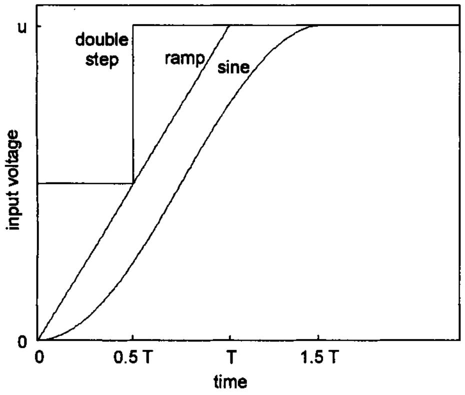


Figure 8a. Input voltage of different control strategies.

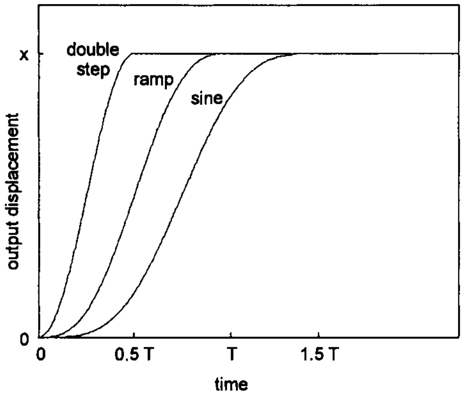


Figure 8b. Response on different control strategies.

For the real system, however, some other facts have to be considered. Because of the piezoelectric actuator being a capacitive load for the power supply, a very high current would be needed to cause a step input for the system. Therefore it is quite difficult to realise the double step input. Using a single ramp as input for a real system with damping causes the system to lag behind the input. This causes significant vibration that can not be reduced satisfactorily. To solve these problems a combination of both strategies can be used: A ramp during the time of *T* causes the system to overshoot and when the system reaches the maximum position, a second ramp keeps the system in position. The time the system needs to reach the equilibrium position with this strategy is slightly longer than *T*. The vibration amplitude of the overshooting system can be reduced to 1/5 what makes the positioning process a lot faster with still being accurate.

Figure 9 shows the excitation and response of the feeding system for one bond cycle. To present the results of the adapted open loop control, the two cases, single ramp excitation (fig. 9a) and adapted double ramp (fig. 9b), are shown.

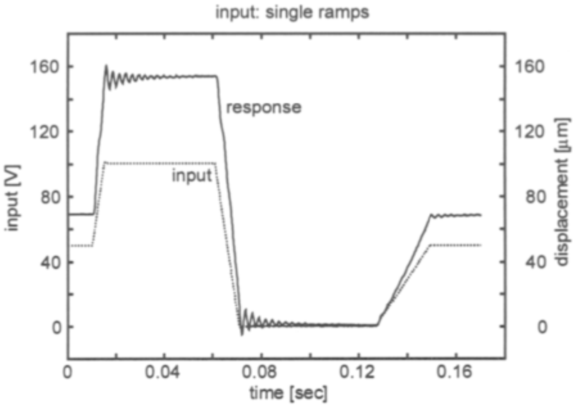


Figure 9a. Response of the wire feeding system during a bond cycle with single ramp input.

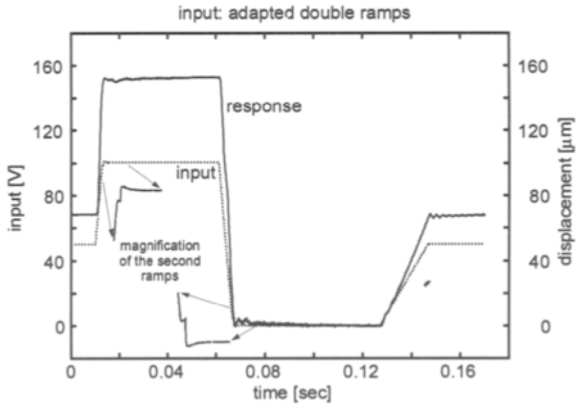


Figure 9b. Response of the wire feeding system during a bond cycle with adapted double ramp input.

An open loop system cannot react on disturbances or parameter variations of the system, but in this special application it doesn't matter. The shift in parameters resulting from the two tasks of the feeding system, tearing off the wire and simply moving it, are insignificant. The stiffness of the system is high enough, that the small force needed to tear off the wire does not alter the systems behaviour significantly. Besides, for the process of tearing off the wire, accuracy is not as important as speed.

Disturbances of the feeding system, occurring from positioning the bond head, are not very strong. Anyway, before touch down of the welding unit, when accuracy is indispensable, the bond head is accelerated very gently, thus disturbances almost vanish.

Should it be important for future applications to reduce even the small effect on the feeding systems accuracy, occurring from moving the bond head, this could be reduced by open loop precontrol. If the transfer function from the input of the positioning axes to the displacement of the feeding system is derived, the information of the input of the axes can be used to actively move the feeding system against the occurring disturbances. However, this is not necessary for the bonding machines presently used.

4. PERFORMANCE OF THE NEW SYSTEM

Compared to the feeding systems currently in use the wire feeding system presented here (fig. 10) has several advantages:

The volume has been reduced to 20 %. This also effects the rest of the bonding machine. The bond head can be designed more compact what significantly reduces its mass. In consequence of that, the whole positioning system for the bond head can be designed lighter and thus a faster positioning can be achieved.

An improvement of the positioning speed of the feeding system could be achieved. With an accuracy of less than 5 % of the stroke length the system is still five times faster than the one currently used.

The handling of the system could be improved significantly. The awkward manual adjustment could be replaced by a new feature: the programmable stroke length. Adjustment of the system has to be done just once and all adaptations of the stroke length can be done by program.

As mentioned before, the system currently in use is strongly affected by wear. Because of the use of flexure hinges and the whole system being one part with no internal relative motion, wear does not occur in significant amount.

5. CONCLUSION

In this report a new wire feeding system for an aluminium wire bonding machine was presented. Compared to the system presently used in bonding machines its performance could be increased significantly. The requirements for the system set by the bonding technology are completely met. There is still potential to make bonding machines faster and more accurate with this new kind of feeding systems.

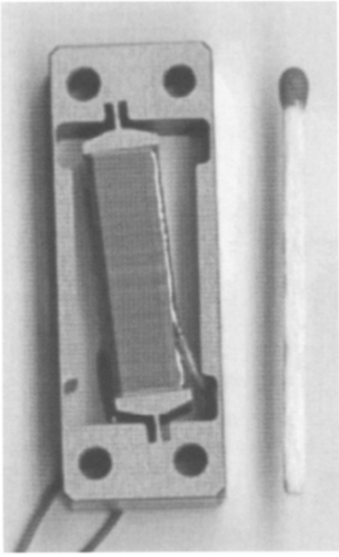


Figure 10. Parallelogram-Actuator: Designed as a wire feeding system for an wedge-wedge-bonding machine.

REFERENCES

1. A. Henke, “Drahthandling - Entwurf eines mikromechatronischen Drahtvorschubsystems für einen Aluminiumdraht-Bonder”, F&M Feinwerktechnik, Mikrotechnik, Mikroelektronik, 5/98, ISSN 0944-1018
2. Deutscher Verband für Schweißtechnik, Technischer Ausschuß, Arbeitsgruppe “Schweißen in Elektronik und Feinwerktechnik”: Drahtbonden, Merkblatt DVS 2810, September 1992
3. T.G. King, ‘Application of Piezoelectric Actuators to Mechatronic Actuation’, Proc. Configuration and Control Aspects of Mechatronics (CCAM), pp 89-105 , Ilmenau, Germany, 1997. ISBN 3-932633-07-5

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Behaviour-based assembly of a family of electric torches

Giovanni C. Pettinaro [a*]

[a]Mechatronics Group, Dept. of Systems Engineering, ABB Corporate Research,
Gideonsbergsgatan 2, S-721 78 Västerås, Sweden

The robot automated industrial task reported in this paper concerns the assembly of a family of electric torches, each made of 6 components: two batteries, a torch tube, a reflector, a glass, and a ring retainer. The robot program accomplishing such a task is developed within the behaviour-based assembly paradigm and it assumes to make use of a two-handed agent equipped only with a differential touch sensor.

The aim of the paper is not to show that the different parts of one torch kit can be assembled together, but to prove that, given a family of assembly kits and knowing how to assemble one, it is possible to program a robot to assemble all of them.

1. INTRODUCTION

The task of mating parts by means of a screwing operation (*e.g.* nuts on bolts) is very common in manufacturing industry[2,7,8]. Such a task is usually automated by means of special machinery employing dedicated tools like electric screwdrivers.

The aim of the research reported in this paper is not to define a substitute for these tools, but indeed to show that it is possible within the behaviour-based assembly paradigm to define a description of the assembly task allowing the manipulator agent to perform not just one particular instantiation of it but the entire class, that is the entire group of assemblies requiring the same sequence of steps (family of assemblies).

1.1. Related Works

The literature concerning screw tightening and involving robots is very little and is most of the time limited to the issues raised by the employment of opportune specialised end-effectors.

An exception is the work reported in [6] which, by developing a dynamic simulation of threaded insertions based on Euler's equations, impulsive forces, and geometric descriptions for threaded parts, also addressed the problem of initial alignment. In this regard, it is worth mentioning the work reported in [3] which shows a technique for determining the alignment of threaded parts using data obtained from minimal force and angular position. Such a technique is based on backspinning a nut with respect to a bolt and then measuring the force spike occurring when the bolt falls into the nut thread. This idea is similar to the one applied in [10], nonetheless the two of them are fundamentally different in the sense that the latter, which is largely inspired on the human beahviour, assumes to perform forwardspinning and does not require any mathematical model in order to start up the thread.

1.2. Architectural Framework

The research reported in this paper is developed within the beahvioural-based approach [5]. Such an architecture may be regarded as a hybrid between the classic totally planning paradigm and the distributed reactive systems [1]. While it shares with the former the characteristic of describing tasks in terms of plans, it does not require a plan to be expressed in low-level details such as specific robot motions. What it indeed requires is a task specification in terms of task achieving units (behaviours) defined as opportune combinations of hardware and software [8].

By decoupling planning and reactiveness, this paradigm frees the planner from the need of deal-

---

[*] The experimental work was carried out at the Department of Artificial Intelligence at University of Edinburgh with a research grant from the Commission of the European Communities (Grant No. ERBCHBICT920012). Current e-mail address: giovanni.pettinaro@secrc.abb.se

ing directly with the real world allowing the agent to work and reason in an ideal one. The interaction with reality is delegated and solved down to the basic bricks of a plan: a limited set of atomic behaviours [8] which, by composition, can define any plan onto their domain of applicability [9].

The execution of a behaviour may lead the system made of agent+environment to different final states which are coded with a number (exit state). There is no success or failure as such when a behaviour is performed. Indeed, each exit state value can be used to let a behaviour reach a different outcome and hence achieve a different goal.

The research described here is carried out within the framework of the paradigm outlined above and is based on the results gathered in [10].

The rest of the paper first describes the experimental testbeds employed and the workcell assumptions laid down (section 2), then it outlines the synthesis of a behavioural plan allowing the assembly of the entire family of the torch kits considered (section 3), then it summarises the experimental results achieved with the aforementioned plan (section 4), and finally, it draws final conclusions (section 5).

2. RESEARCH ASSUMPTIONS

The electric torch chosen as the experimental testbed in this work is composed of six parts: a tube, a reflector with a bulb, a plastic glass, a threaded large ring to retain the reflector and the glass, and two batteries which I labeled A and B (figure 1). The reflector with a bulb is actually a subassembly made of the reflector itself and an appropriate bulb with a threaded end which is used to tighten the bulb to the main body of the reflector. However, since the aim of this paper is not simply to show that the different parts of a particular kit can be composed together, but it is indeed to show that, given a family of similar kits, the same behaviour-based assembly plan is enough to allow a behavioural agent to assemble all of them without further information, I chose for a question of convenience to consider it as a whole part.

The torch kit described above is one of three models (figure 2). Since each of them requires the same sequence of steps in order to be assembled, they belong to the same family of products.

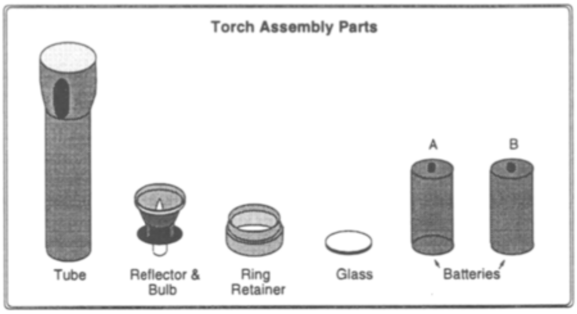


Figure 1. Torch assembly Kit.

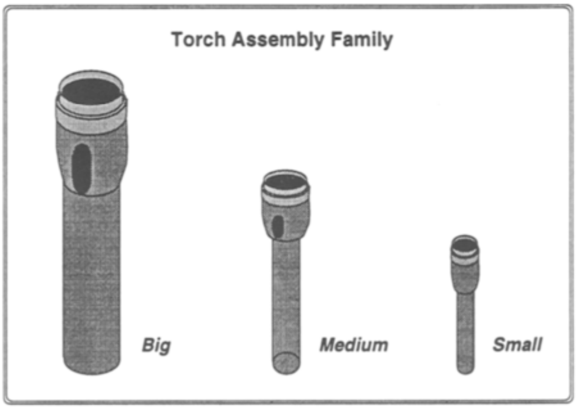


Figure 2. Torch Family.

As regards the workcell, I assumed to have a very standard set-up: a Cartesian SCARA manipulator with 5 d.o.f. (*Adept* 1) programmed in VAL II [10] and a teach pendant for storing workcell locations (a 6 dimensional vector made of 3 components for x, y, z, and 3 components for pitch, yaw, and roll angles of the wrist). Notice that the sixth component of a location is available but not used.

As regards the end-effector, in order to be as general as possible, I assumed to employ a two-fingered jaw gripper [11] electrically driven directly by the robot controller.

As far as sensor are concerned, I assumed to use a very simple haptic sensor made of a piezo-film and wrapped around the fingertips of each phalanx [4]. Such a film has the characteristic of producing an event signature whenever a change in its shape occurs, which means that such a device triggers a signal upon either contact or release of an object. Given its differential nature, it cannot read forces or torques, nonetheless it is still quite enough to cope with the screwing tasks involved in the assembly task considered.

As regards the fixture, I have to point out that, as I initially verified, the set-up described above is sufficient with suitable jigs to make the agent accomplish the assembly task. However, I noticed that this did not add any valuable generality to the results of the research. Thus, in order to simplify the work, I thereafter assumed to rely on a second two-fingered jaw gripper fixed to the table and used as a mere programmable vice directly driven by the robot controller. The system I ended up with (two equal robot controlled grippers) might be regarded to a certain extent as a sort of two-handed robot [12].

3. BEHAVIOURAL PLAN

Examining the kit described in section 2 (cf. figure 1), any human operator would carry out the following sequence of steps in order to carry it out:

1. inserting each battery inside the tube,
2. placing the reflector on top of the tube,
3. placing the glass on top of the stack, and finally
4. tightening the retainer to the tube.

Such a description (cf. figure 3 for a graphical representation) is sufficient for assembling not just one particular kit but an entire family of them. Holding this decomposition as a model for the assembly plan, I have to point out that step 1,2,and 3 are to be regarded in each case as stacking tasks, that is as sequences of monitored point-to-point robot motions, whereas step 4 is instead a form of screw fastening task.

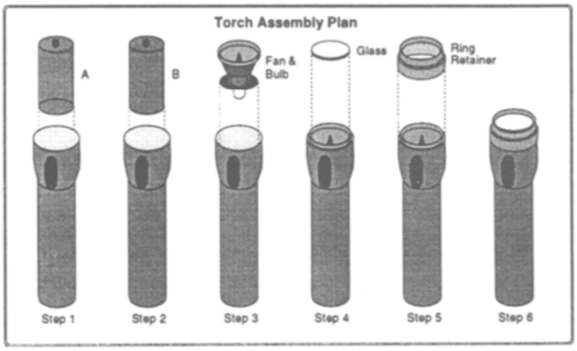


Figure 3. Diagram for describing a Torch Assembly.

The plan outlined above is a very good description of the assembly task to be performed easily understandable by any human operator. The fact that at the bottom of the tube there is a spring pushing up the batteries and everything else stacked on top of them does not affect its descriptive power: any human thanks to his powerful compliant hands can still overcome the difficulty of mating the different parts under the pressure of the spring. Unfortunately, robot agents are not as flexible: once a part is stacked and released on the torch tube, the pressure of the spring pushes the part slightly away from that position introducing extra uncertainty which the robot can not sense.

In order to get around such a problem, I decided to divide the assembly task in two independent subassemblies which can be reliably built in separate stages (figure 4):

- batteries insertion in the torch tube, and
- mating of reflector with its matching glass and ring retainer.

As can be easily gathered by figure 4, the two aforementioned subassemblies require similar pick-and-place operations.

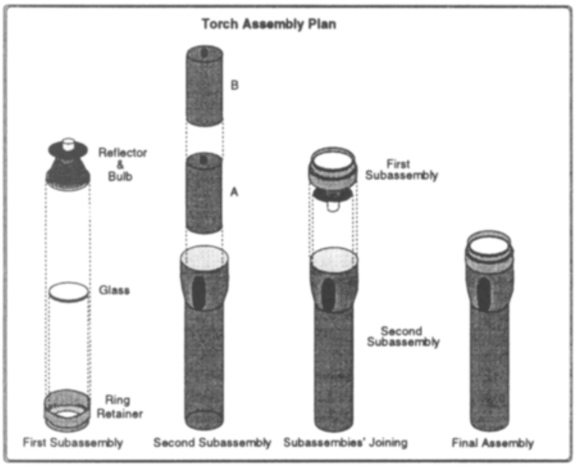

Torch Assembly Plan
B
A
Reflector & Bulb
Glass
Ring Retainer
First Subassembly
Second Subassembly
First Subassembly
Second Subassembly
Subassemblies' Joining
Final Assembly

Figure 4. Torch Subassemblies.

By relying on previously developed picking and stacking basic behaviours (***pick-up*** and ***stack***, respectively), each pick-and-place can be defined as reported in the flow chart of figure 5 and instantiated with four input parameters: a textual command message stating which part has to be collected, the location of such a part, another textual command message stating where the part held has to be released, and finally the location of the stack.

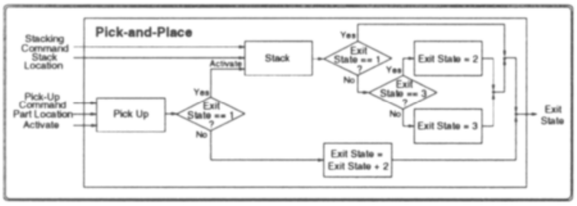

Pick-and-Place
Stacking Command
Stack Location
Pick-Up Command
Part Location
Activate
Pick Up
Yes
Exit State == 1 ?
No
Activate
Stack
Yes
Exit State == 1 ?
No
Yes
Exit State == 3 ?
No
Exit State = 2
Exit State = 3
Exit State = Exit State + 2
Exit State

Figure 5. Pick-and-Place Flow Chart.

Since both subassemblies involve the collection and stacking of two parts onto a third one, each subassembly may be modeled with a composition of three pick-and-place behaviours: one for placing the base, and two for stacking the other two parts.

Notice that I assume each subassembly to be carried out at the same site (table gripper location). Thus, as soon as the first one is successfully completed, it is necessary to move it away to some known work-cell location in order to make room for the other. Once this last is accomplished, next stage is joining it with the former. In this regard, notice that the torch tube is still firmly held by the table gripper at the end of the second subassembly. Thus, the joining process can be carried out by keeping the tube clamped with the table gripper and by going with the right one first to fetch the other subassembly back from where it was temporarily laid down (temporary site) and then to fasten it on top of the tube (cf. figure 6), after, of course, having opportunely re-oriented it so that to match the threads of the retainer and the tube.

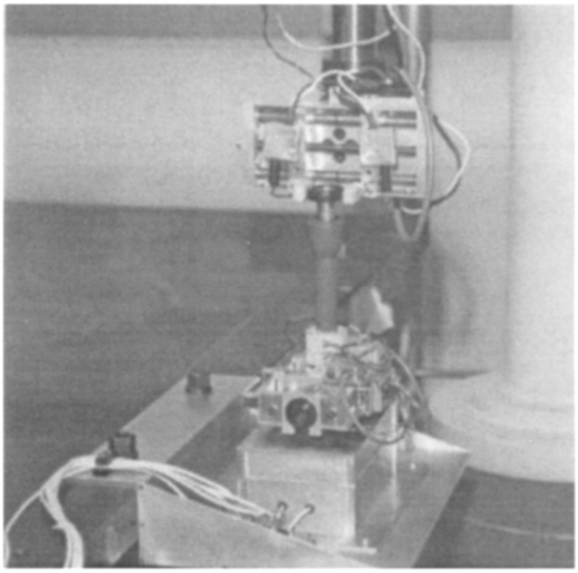


Figure 6. Screw Fastening of the Two Torch Subassemblies.

Summarizing the discussion above, the plan to assemble a torch kit can be divided in the following five stages:

- mating of ring retainer with glass and reflector (subassembly 1),
- relocation of the first subassembly to a temporary site,
- mating of torch tube with the two batteries (subassembly 2),
- retrieval of first subassembly from its temporary site and its reorientation and stacking on the torch tube,
- tightening of the first subassembly to the second one.

4. ASSEMBLY EXPERIMENTS

As pointed out earlier, since the goal I aimed to reach was to develop a behavioural program capable of achieving the assembly of a family of three different torches, I set three different type of experiments: one for each kit available (large, medium, and small).

The first one involves the largest kit. To begin with, I placed the six parts of the kit (ring retainer, glass, reflector, torch tube, battery 1, and battery 2) at specific locations within the work-cell. Then, by employing the robot teach pendant, I stored them as their home locations. Similarly, I stored the locations of both table gripper and temporary site. Once this set-up stage was completed, I implemented the 5 step plan reported at the end of section 3 by using the *pick-up*, *stack*, and *screw* modules available in the library of the behavioural agent used. The first four of them did not show any particular problem. Indeed, testing them separately by means of one sequence of 60 runs, I recorded the full success in each case (100% success rate).

As regards last step (screw fastening of the ring retainer to the torch tube), I recorded a slightly different result: 57 successful fastenings out of 60 trials (95% success rate). The three failures (5% failure rate) were all due to a wrong start of the screw thread caused by the spring at the bottom of the tube which indirectly pushed the reflector out of axis. It is interesting to point out that all the failures occurred within the first 20 trials. The explanation of such an outcome lies in the wearing off of the spring which, after many repeated experimental trials, started losing part of its elastic force.

After having tested all the different steps, I ran the whole behavioural plan in a sequence of 60 complete assemblies of the torch kit recording 58 successful ones (*ca.* 97% success rate). The two failures were again due to the spring push.

The second experiment described involves the medium sized torch kit. Following the same procedure carried out for the largest kit, I placed the different parts at specific sites and recorded them as their home locations. Then, I took the same assembly program used before and I instantiated it with the new home locations keeping the other input parameters (table gripper and temporary site locations, and pickup and stacking commands) unchanged.

In order to test the behavioural plan, I repeated the same pattern of tests outlined earlier, that is a sequence of 60 complete assemblies, where I recorded 59 successes (ca. 98% success rate). The only failure was due also in this case to a wrong start of the screw thread caused by the spring push. The higher success rate with respect to the previous assembly kit is explained by the small physical size of the spring which, because of its smaller coefficient of elasticity, generates smaller forces opposing the screw fastening.

The last experiment reported concerns the smallest sized torch kit. After having placed and stored the home locations of the new parts involved, I instantiated again the same behavioural plan by using these new locations and by keeping the other input parameters unchanged.

The result gathered after a sequence of 60 runs was 60 successful assemblies (100% success rate). The higher success rate with respect to the two previous kits is mainly due to the very small size of the spring involved which is capable of producing only very tiny push disturbing the successful screw fastening.

The results obtained by the three experiments are summarised in table 1.

Table 1
Torch Kit Experimental Results.

Kit Type	Successes	Rate
Largest	58/60	97%
Medium	59/60	98%
Smallest	60/60	100%

5. CONCLUSIONS

The behavioural torch assembly plan presented in this paper was loosely modeled on the human-like decomposition of such a task and its development was carried out within the behaviour-based assembly approach. The work assumed to rely on a library of behavioural units, defined in terms of a limited set of basic ones.

The behavioural plan implementing the aforementioned decomposition showed a considerable reliability and robustness as well as a very appealing generality within the same family of assemblies which points out the potential expressiveness offered by the behaviour-based assembly paradigm.

REFERENCES

1. Rodney Allen Brooks. A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation*, 1(2):14–23, April 1986.
2. Carlton Byrne. A Literature Survey of Classification Systems for Assembly Processes. Cardiff Technical Note ARC n° 17, Department of Mechanical and Manufacturing Systems Engineering, University of Wales Institute of Science and Technology, October 1987.
3. M. A. Diftler and I. D. Walker. Determining Alignment between Threaded Parts Using Force and Position Data from a Robot Hand. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, pages 1503–1510, Albuquerque, New Mexico, U.S.A., 20th-25th April 1997.
4. Taehee Kim, Chris A. Malcolm, and John Hallam. Developing of Vibration Sensors as Event Signature Sensors in Assembly. In *Proceedings of the International Conference on Robotics and Manufacturing*, pages 39–41, Oxford, U.K., September 1993.
5. Chris Malcolm and Tim Smithers. Programming Robotic Assembly in Terms of Task Achieving Behavioural Modules. *Structural Programming*, 1988.
6. Edward J. Nicolson and Ronald S. Fearing. Dynamic Modeling of a Part Mating Problem: Threaded Fastener Insertion. In *Proceedings of the 1991 IEEE International Workshop on Intelligent Robots and Systems (IROS '91)*, pages 30–37, Osaka, Japan, 3rd-5th November 1991.
7. Giovanni C. Pettinaro. *Basic Set of Behaviour for Programming Assembly Robots.* PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.
8. Giovanni C. Pettinaro. Basic Set of Behaviours for Programming Assembly Robots. In *Robotikdagarna 1997*, pages 61–72, Linköpings Tekniska Högskolan, 10th-11th June 1997.
9. Giovanni C. Pettinaro. Definition of a Behaviour-Based Language for Industrial Robots. In *Proceedings of the International Scientific Conference on Artificial Intelligence in Industry*, pages 237–246, Stara Lesna, High Tatras, Slovakia, 22nd-24th April 1998.
10. Giovanni C. Pettinaro. Human-Modeled Robot Screw Fastening. In *Proceedings of the 1998 IEEE International Conference on Intelligent Robots and Systems (IROS '98)*, Victoria Conference Centre, Victoria, B.C., Canada, 12nd-16th October 1998.
11. Giovanni C. Pettinaro and Chris Malcolm. Electric Gripper Development. D.A.I. Technical Report n° 28, Department of Artificial Intelligence, University of Edinburgh, September 1994.
12. Giovanni C. Pettinaro and Chris Malcolm. Two-Handed Robotic System. D.A.I. Technical Report n° 34, Department of Artificial Intelligence, University of Edinburgh, May 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Electronic Reconfiguration in Flexible Winding Type Machines

N Jakeman (1), WA Bullough (1), PH Mellor (2)

SMMART Unit (1) & EMD Group (2), The University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.

ABSTRACT

Mechanisms, which allow a change of duty to be achieved by pre programmed electronic switching (alone), are reviewed against a demonstrator task of filament package winding. The ultimate limit of performance is set by strain, the precision of the sweep requirement and availability of materials. The review will include conventional stepper/servo motor limitations, catch methods based on electrostrictive and electromagnetic stimulation and more novel electro/magneto-rheological designs plus a tubular linear actuator with Halbach array. Prospects will be related to said material properties, thermal considerations and/or the demagnetisation properties of any permanent magnets used in the design.

1 INTRODUCTION

Industry is always looking to develop new processes and machinery so that the quality of its output can be improved and the cost of its operation reduced.

To facilitate this, two different areas of design are prominent, the speed of operation and the flexibility of control of manufacturing machinery. Reconfiguration of current manufacturing machinery often involves the replacement of linkage mechanisms forming the mechanical advantage and velocity ratio magnifier. Furthermore, under highly demanding output profiles the materials from which these mechanisms are made may well reach an unpredictable manufactured fatigue limit.

Therefore a new genre of machine is investigated which addresses the speed of response of a machine to a change in the operational requirements whilst still offering an adequate supply of force, velocity and displacement [1]. The machine should also have the capability to be reconfigured without geometric alteration in order to eliminate downtime and to avoid stress concentration

The switching of a drive in fixed geometry conditions may be considered analogous to the stepping action featured by that of a digital electronic waveform generator. Figure 1 illustrates how the rapid switching may be used to generate a waveform.

The motion profile illustrated in figure 2 is considered as a suitable demonstrator to the concept. Position variability in the traversing mechanism used to guide yarn onto packages used in the textile industry is required to eliminate "bundling" at both ends of the packages.

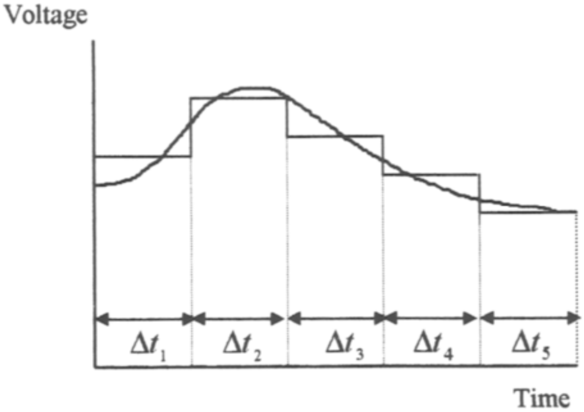


Figure 1: Busbar catching analogy for flexible drive methodology.

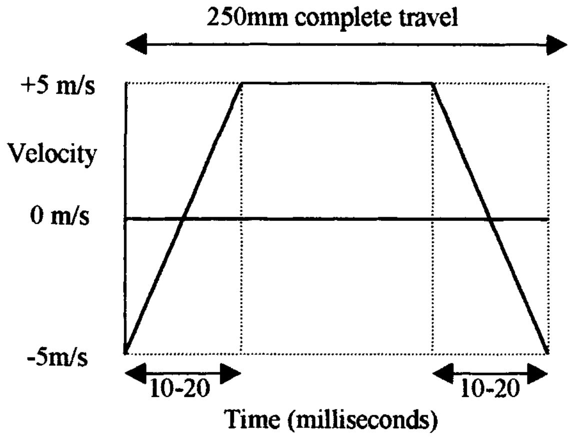


Figure 2: Demonstrator velocity profile.

This clearly shows the need for a machine of the type described above. Accelerations and position accuracy of 50-100g and ~0.25mm are desired respectively.

The flexibility of existing drive solutions is primarily limited by their poor acceleration capability caused as a result of the large electrical and mechanical time constants associated with their design. As a result, various problems are commonly encountered when trying to achieve demanding motion profiles. Problems that include the transfer of heat away from the drive mechanism and the demagnetisation of any permanent magnets that may have been incorporated into the design of the drive.

However, the ultimate limitation placed on dynamic performance is that of the material characteristics of the velocity ratio magnifier. Consider the material of length l, cross sectional area A, and density ρ illustrated in figure 3.

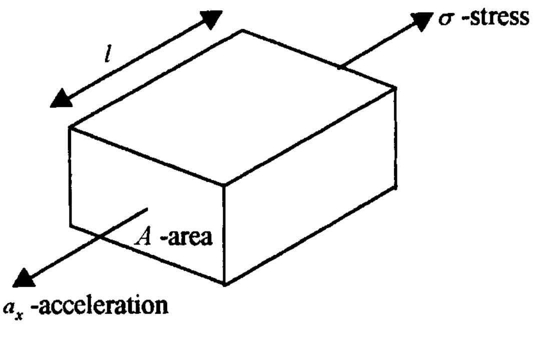


Figure 3: Model used to analyse the material limitation posed by the velocity ratio magnifier.

The force required to accelerate the material at a rate a_x, is given by:

$$\sigma A = \rho A l a_x \tag{1.1}$$

Where σ is the stress exerted on the material.

The material is required to illustrate elastic properties, i.e. it must obey Hookes law, therefore:

$$E = \sigma/\varepsilon \tag{1.2}$$

Hence:

$$\frac{E}{\rho} = \frac{la_x}{\delta/l} = \frac{l^2 a_x}{\delta} \tag{1.3}$$

Where δ is the extension of the material, which will ultimately limit the precision of position.

An effective belt of length 250mm is considered suitable for our demonstrator. Therefore, if we consider a precision of position of 0.1 mm, set by a maximum strain of 0.04% in the belt, for accelerations of 1000g, a value of E/ρ is obtained as $6.25\times10^6\ m^2s^{-2}$. Comparing this to E/ρ for all materials shows that this is possible but only a few materials, notably Kevlar, would be suitable for the belt [2]. Hence an acceleration of 100g is set as a more realistic limit, whilst a strain of 0.1% in the belt would allow a precision of ± 0.25mm.

Furthermore, for this target to be achieved using an electrical drive solution, it must have a sufficiently low electrical time constant, preferably in the order of one millisecond. What follows is a linear drive methodology illustrating the limitations placed on the performance of both commercially available drive technologies and more novel emerging actuation techniques.

2 LINEAR DRIVE METHODOLOGY

The review is carried out under three different categories commencing with a simple stepper motor, this being an obvious solution to applications requiring small position increments. Various clutching methods incorporating both electromagnetic actuators and smart materials are also considered as a means of coupling a high inertia input to a low inertia load. Finally, servo drives are considered; servo drive being defined herein as any electromechanical motor/actuator which, when encoded and set up in a closed loop configuration with position control, can be used to drive a load through a required velocity profile.

2.1 The Stepper Motor

The stepper motor provides a simple, cost-effective solution to many low speed applications requiring position increments. The device can be driven open loop without position feedback thus eliminating the need for an expensive encoder and hence reducing the complexity of the drive and controller.

However, on performance the stepper motor simply cannot provide the desired criteria for good flexibility to be achieved. At high step rates, typically 200-300 steps per second, the motor begins to operate in its characteristic "pull out" torque region where rated torque simply cannot be achieved. Hence beyond this speed the torque/inertia ratio deteriorates. The large electrical time delay associated with the motor (L/R), typically 10 milliseconds, also restricts its response time.

Finally, the fixed step angle results in a lack of flexibility in the step resolution. For example a 200 step motor (1.8^0 per step) used as our demonstrator (paper exercise) could achieve no more than a ± 1.25mm accuracy of position. Micro stepping would improve this, however this technique is difficult to implement due to the problem of torque production and repeatability.

2.2 Catch Mechanisms Incorporating Electromagnetic Actuators and Smart Materials

One possible method of achieving rapid acceleration of a light load is to latch it onto the rotor of a motor whose inertia is considerably greater than that of the load.

Figure 4 illustrates how two clutches, A and B, could be used to latch two separate motors, driven in opposing directions. Reciprocating linear motion is derived on the belt (G) coupling the two pulleys (F) by switching the two clutches alternatively in a time pattern. Various profiles can be achieved without the need for geometric alteration of the machine simply by altering the frequency at which the catches are controlled.

Electro-rheological fluid (D) is used in this example to couple the input rotor to the output rotor of each clutch, C and E respectively. However it is important to review all of the latching methods considered appropriate for use on the demonstrator in order to find ultimate limitations on their performance.

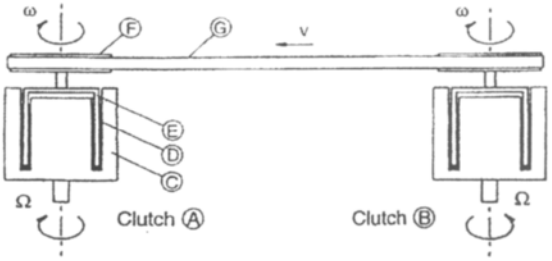


Figure 4: Linear drive using two catch mechanisms

The performance of a simple electromagnetic clutch is limited solely by the mass of the clutch plates and the large L/R electrical time constant associated with the windings. Figure 5 illustrates the typical torque response of an electromagnetic clutch after the application of a signal to its windings at t=0 [3].

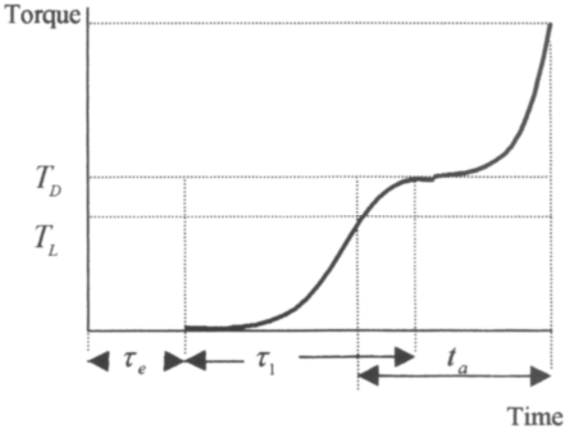


Figure 5: Torque profile for an electromagnetic clutch

The engagement time is equal to the sum of the electrical time constant, τ_e, and the dynamic torque rise time, τ_1. The later being defined as the time taken for the torque to rise to it's rated dynamic value, T_D. Typically the engagement time would lie in the range 15-30 milliseconds. If field forcing techniques were used to reduce τ_e, this figure could be improved. In addition to this a finite time will elapse, t_a, whilst acceleration of the output member takes place. This occurs after the load torque, T_L, has been established and is calculated as[3]:

$$t_a = J(\omega_2 - \omega_1)/(T_D - T_L) \qquad (2.1)$$

Where J is the combined inertia of the clutch plates and load and $\omega_2 - \omega_1$ is the final velocity minus the initial velocity. At best, accelerations of 20-25g are considered possible using this technology for our demonstrator task.

The electrorheological ER fluid clutches used in figure 4, potentially offer a significantly faster engagement time than that of an electromagnetic equivalent. Whereas the latter requires an iron circuit to focus the magnetic field, the ER fluid clutch requires only a pair of electrodes, one placed either side of the fluid, in order to generate the large electric fields required in the fluid, typically 7 kV/mm. As a result the device has an electrical time constant, τ_e^*, in the order of microseconds due to its RC nature, where C is the effective fluid capacitance.

The electron hydraulic time delay, τ_m^*, is defined as the time taken for the fluid to generate full torque after τ_e^* has taken place. For a typical concentric clutch of the type illustrated in figure 4, τ_m^* dominates the engagement time and is typically in the order of one millisecond [4]. Therefore the device has the potential to respond rapidly to a change in operational requirements. After τ_e^* has taken place, the clutch will ideally run up to 99% of its final velocity in a time t_0, given by:

$$t_0 = 4\rho h^2/\mu \qquad (2.2)$$

Where ρ is the density of the fluid, h is the gap width and μ the zero volts viscosity. For a typical concentric ER fluid clutch with h=0.5mm, ρ=1000 Kgm^{-3} and μ=0.1 Pa, a value of t_0 is obtained as 10 milliseconds. Therefore, accelerations in the order of 50g can be potentially realised for our demonstrator task.

Reducing the gap width, h, could improve this, however the drag on the faces, T_d, of the non energised concentric clutch would increase significantly, as:

$$T_d = 2\pi r^3 \omega l \mu / h \qquad (2.3)$$

Where r is the radius of the clutch and l the length of the clutch plates. Therefore the performance of the clutch becomes limited by the need for the effective transfer of heat away from the device. With appropriate cooling the upper limit on performance would be dictated by the 0.1% strain in the belt discussed earlier. Accelerations greater than 100g could therefore be realised if an internally stable fluid of strength ~10kPa were available.

Results obtained using ER fluid clutches on the demonstrator rig have suggested that accelerations in the order of 50g are possible, however this has been achieved but with poor reliability and evidence of avalanche current breakdown within the fluid at high duty loadings has been recorded [5].

Magnetorheological (MR) fluid could be used in the catch mechanism in order to generate significant yield stress in the order of 50-100kPa [6]. However, the need for an additional iron circuit to focus the magnetic field introduces a significant electrical time delay in the device of an L/R nature, typically in the order of 10-12 milliseconds [6]. Field forcing techniques can be used to reduce this or alternatively a method of controlling the fluid using two permanent magnets could be sought.

The run up time, t_0, given in equation 2.2 is considerably greater for an MR device than that of its ER equivalent due to the greater density of the fluid. For a gap size of 0.5mm and fluid density of 4000 Kgm^{-3} a typical value for t_0 would be in the order of 20 milliseconds. Thus emphasising the need to reduce the gap size in order to improve the acceleration capability of the device.

In doing this, the drag on the non-energised clutch plates is increased even more dramatically than that observed when increasing h in an ER fluid device. This is due to the greater zero volts fluid viscosity, typically 0.2-0.3 Pas [7], associated with MR fluid. Once again the performance of a MR fluid latch is limited by the need to transfer heat away from the device.

Piezo electric actuators are capable of generating large pressures in a significantly small period of time, typically 4000 N/cm^2 of piezo stack, in less than 100 μ seconds [8]. However their suitability for use in demanding applications is limited by two factors.

Multilayer piezo ceramic stacks are susceptible to any force that may be acting in a direction differing to that in which the device generates force itself. Hence any shear that may provide a threat must be eliminated from the stack if fracture is to be avoided. Secondly, piezo actuators are capable of providing only small displacements. Figure 6 illustrates how a linear drive could be implemented using a piezo actuator.

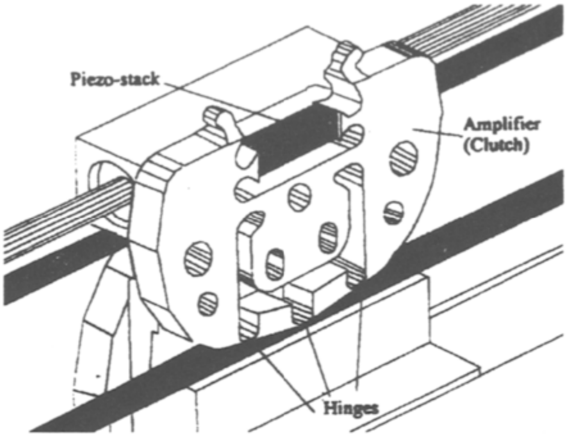


Figure 6: Example of a piezo actuated linear drive.

The device uses a flexure hinge to amplify the displacement of the piezo stack, typically from 15 μm to 110 μm [9], in order to latch the slider mechanism to a moving strip. The arrangement provides sufficient displacement for successful latching to be achieved whilst eliminating shear forces from the stack. However, the mass of the flexure hinge and slider arrangement combined with the loss in output force due to displacement amplification prevents rapid acceleration.

Alternatively a piezo stack of sufficient length, say 10cm, could be used to latch a radial clutch. This would provide a displacement of 0.1mm without the need for amplification. The drag on the faces of a radial clutch is given by:

$$T_d = \frac{\pi}{2} r^4 \mu \frac{\omega}{h} \tag{2.4}$$

Where h is the distance between the two clutch plates. For a clutch sized to produce a drive stress of 2 MPa using a piezo stack of 1 cm^2 cross sectional area, a drag of 57 μ Nm would be experienced if the clutch were run dry and 0.62 Nm if run wet. Given that the clutch can generate some 100 Nm, drag would no longer be of concern. However, eliminating shear from the stack could prove more troublesome. Consideration must also be given to the transfer of heat away from the surface of the stack caused as a result of the rapid switching of the device (18 Hz required).

2.3 Servo Drives

An a.c. motor operated under closed loop position control could provide an alternative method of providing high performance, high bandwidth flexible motion for light load conditions. However, both the induction motor and switch reluctance motor can be rejected immediately as potential solutions. Despite a considerable improvement in the performance of the induction motor due to the development of vector control, dynamic performance is still restricted by the large rotor time constants associated with its design. Meanwhile torque ripple (as high as 20%), acoustic noise and poor velocity control bandwidth continue to affect the performance of the switch reluctance motor.

The availability of high performance brushless servo drives has occurred due to continued development of high remenance rare earth magnets, with high coercivity and consequently good resistance to demagnetisation. These machines have low electrical time constants and good Torque/Inertia capabilities giving rise to a motor drive with a "stiff" response to changes in operational requirements. Figure 7 illustrates the response of a typical commercial brushless motor to a step input.

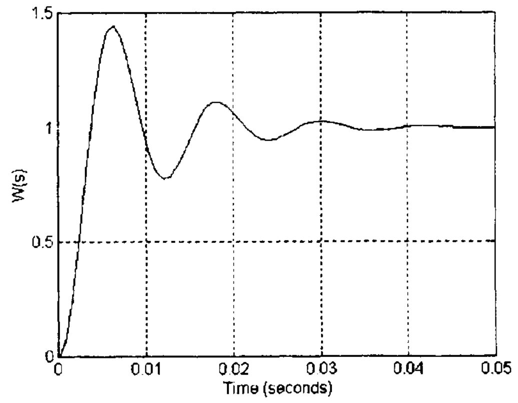


Figure 7: Response of a brushless servo motor to a step input.

In this particular example, the motor has been set up with an outer velocity control loop and inner current loop with PID (Proportional, Integral, and Derivative) control structure. Used for our demonstrator task, this motor can provide the required 100g acceleration whilst phase and gain margins of 51.3^0 and 40dB suggest that good position control can be realised. In performing the profile illustrated in figure 2 the motor does not exceed its rms rated torque value of 2.4Nm, thus overheating of the windings and demagnetisation of the permanent magnets are not an issue.

A recent development in brushless servo design involves the use of a different type of magnet array than that conventionally used in machine design. The Halbach array, illustrated in figure 8, displays the following advantages:

(1) A fundamental field some 1.4 times stronger than that of a conventional array.
(2) The return field path is confined within the magnets, thus eliminating the need for any back iron.
(3) The near perfect sinusoidal magnetic field distribution yields benefits of reduced cogging torque, whilst the machine can be operated using simple a.c control structures.

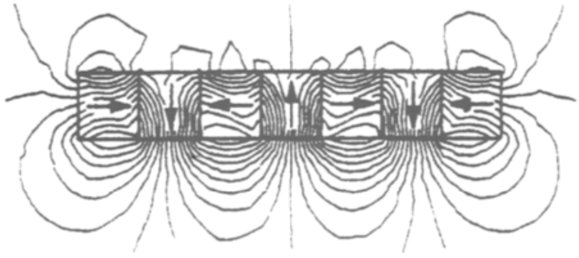


Figure 8: The Halbach Array.

The use of Halbach arrays leads to machine designs with low inertia external rotors by bonding the array to a low density material or composite such as CFRP. One such machine, designed for a powered roller application, is described in ref [10]. Accelerations at the surface of the roller of 11g were recorded using a roller of inertia $1\times10^{-2}\,Kgm^2$.

The technology could be incorporated into a small air cored linear actuator in order to eliminate the need for a velocity ratio magnifier. The provision of a cooling path to transfer heat away from the windings allows the electric loading, Q, of the machine to be increased. Consequently the shear stress, σ, acting on the surface of the rotor is also increased as:

$$\sigma = B.Q \tag{2.5}$$

The magnetic loading, B, is dependant on the remenance of the magnets and the topology of the machine. Therefore the performance of the actuator will be limited by the availability of magnets with sufficient magnetic remenance and the maximum electric loading that can be tolerated before either overheating of the windings and/or demagnetisation of the magnet array takes place. Figure 9 illustrates the design of a two-phase tubular linear actuator incorporating Halbach array [11].

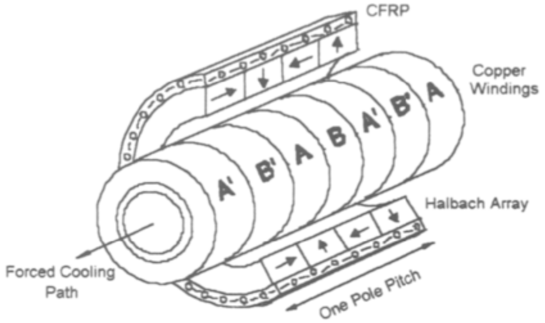


Figure 9: Tubular linear actuator incorporating Halbach array.

Assuming that all of the heat generated in the windings is transferred via conduction into the central cooling path and that the air surrounding the windings acts as a perfect insulator, the temperature, θ_m, at the external surface of the windings can be found as:

$$\theta_m = \frac{q_v}{k}\left[\frac{r_i^2 - r_o^2}{4} + \frac{r_o^2(\ln r_o - \ln r_i)}{2}\right] + \theta_c \tag{2.6}$$

Where k is the thermal conductivity of the windings, q_v the heat generation density in the windings and θ_c the temperature at the internal surface of the aluminium tube. r_i and r_0 are the respective inner and outer radii of the windings.

This thermal model has been used to calculate the heat generated in the windings of a linear actuator performing the demonstrator velocity profile. The linear actuator was designed with an internal stator radius of 10mm, an airgap of 0.5mm and a magnet thickness of 1.4mm. The latter being sized by the need to provide an adequate thickness of magnet to provide sufficient withstand to demagnetisation.

Using 2.6 and equations for current density given in [11], an optimal relationship between motor pitch and external winding radius was discovered which can be used to minimise the heat generated in the windings for a given winding thickness (Figure 10). The analysis was based on a machine design that used sintered NdFeB permanent magnets of 1.1 Tesla remenance.

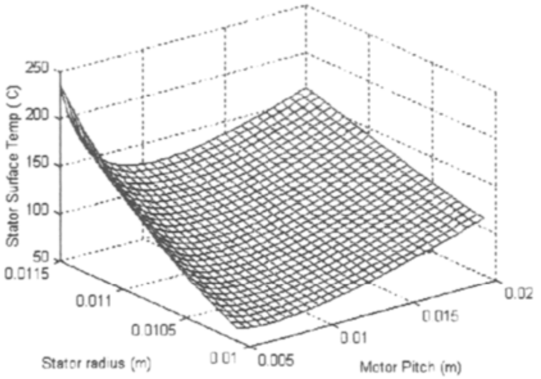


Figure 10: Relationship between motor pitch, winding radius and stator surface temperature.

Beyond external radii of 11.3mm the magnets will be prone to demagnetisation. However, at this point a value for the magnet pitch is obtained as 10.5mm, a value for which manufacture of the magnets themselves would become very difficult. One potential solution to overcome the need for cutting block magnets is to impregnate the CFRP with permanent magnet particles, however the required magnetic flux density of 1.1 Tesla could no longer be obtained using this approach.

3 CONCLUSIONS

A linear drive methodology has been presented for assessing the possibilities of providing high performance flexible motion by electronic switching of a drive in fixed geometry devices for light load conditions. The review suggests that various drive technologies have the potential to meet the criteria defined for high speed flexible machines defined here in. However, in order to achieve this further work is required in several areas.

The phenomenal force capability of piezo electric actuators can only be realised if a method of eliminating shear on the piezo stack can be found. Alternatively a method of strengthening the stack could be sought.

In order to take advantage of the rapid response times inherent with ER fluids, a stable fluid is required which can provide 10kPa yield stress. Furthermore a method of cooling such a device must be found.

Brushless servo motors already provide a feasible option, however their complexity and cost make them somewhat unattractive to manufacturers requiring several hundred independent machines. Therefore the full potential offered by machines, such as those which incorporate Halbach arrays of permanent magnets, in terms of reduced complexity and improved performance must be explored.

4 ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Professor T. King for his contribution to the review and the EPSRC for funding the research.

5 REFERENCES

[1] Bullough WA; The Third Wave of Machines; Endeavour; Vol 20, No 2; June 1996; p50-55.

[2] Yates JR, Lau D, Bullough WA; Inertial Materials for Flexible, High Speed Machines: Perspective, Review, Future Requirements; Proc. Actuator 94, conf. Bremen; p275-278.

[3] ZF Electromagnetic clutch catalogue.

[4] Bullough WA, Johnson AR, Hosseini-Sianaki A, Makin J, Firoozian R; The Electro-Rheological Clutch: Design, Performance Characteristics and Operation; Proc. IMechE Part 1; Vol 207, No 2; 1993; p87-95.

[5] Bullough WA, Johnson AR, Tozer R, Makin J; Methodology, Performance and Problems in ER clutch based Positioning Mechanisms; Proc. International conference on ERF/MRS and their applications, July 1997; Yonezawa; at press World Scientific, Singapore.

[6] Ginder JM, Sproston JL; The Performance of Field Controllable Fluids and Devices; Proc. Actuator 96; 26-28 June 1996; Bremen, Germany.

[7] Rheonetic Magnetic Fluid Systems data.

[8] Thornley JK, Preston ME, King TG; A Fast Electro-Mechanical Clutch Element Using A Piezoelectric Multilayer Actuator; Colloqium on Robot Actuators, IEE; October 1991; London; p1-4.

[9] King TG, Thornley J, Xu W; Piezoceramic Actuators for Mechatronic Applications; Proc Tampere Int Conference on Machine Automation, February 15-18 1994; p569-583.

[10] Atallah K, Howe D, Mellor PH; Design and Analysis of Multi Pole Halbach Cylinder Brushless Permanent Magnet Machines; Proc. EMD 97.

[11] Kim WJ, Berhan MT, Trumper DL, Lang JH; Analysis and Implementation of a Tubular Motor with Halbach Magnet Array; Proc. IEEE; 1996.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronic Design of a Leather Spray System

Sherman Y. T. Lang, K. P. Rao, Lai To Ching
Department of Manufacturing Engineering and Engineering Management,
City University of Hong Kong, Kowloon Tong, Hong Kong

Leather skins require a variety of processing steps. Many leather products require coatings such as paints and lacquer to enhance appearance, protect the surface and mask minor visual defects. The coatings are often applied by spraying. For large production volumes, automated machinery is readily available. For small factories finishing leathers in Hong Kong, large production lines are too costly and bulky. Most small leather processors spray coatings manually requiring skilled workers to produce high quality products. Such skilled workers are in short supply due to the potential health risks of the job. A system to automate the spraying of coatings on leathers has been designed and a prototype system implemented to assess the feasibility. A safe, compact machine, adaptable to the factory environment and space available giving good spraying quality, having low maintenance requirements, capable of manual control as well as automatic spraying was required. The following design features were achieved in the prototype system: selectable automatic and manual operation, remote control, explosion safe, paint conservation, adjustable spray nozzle height, fast spraying operation, high quality spraying, compact size, safety, low cost, reliability, and energy conservation.

1. INTRODUCTION

Most leathers require coatings such as paints to enhance appearance, protect surfaces and compensate for small defects and blemishes. Coatings are commonly applied by spraying. Manually spraying requires a skilled worker to produce a high quality products. Despite the use of ventilation and filtration systems, the spraying process produces a dirty and hazardous working environment. Suspended particulate can cause respiratory disease and solvents in lacquer based coatings are possible carcinogens. Without proper ventilation there is also a danger of explosion and fire. Due to the nature of the occupation, skilled workers are in short supply and employee turnover is high.

Although a variety of automatic spraying machines can perform different spraying jobs, they are not suitable for the local leather industry due to size and cost. Initial capital and running costs are very high for small and medium sized factories that comprise the bulk of local leather industry. This is mainly because they are designed for larger production runs than is the norm in the local industry. The automatic machines usually require large floor spaces which most local factories cannot provide.

To address the needs of the local industry, the design and development of an automatic leather spraying machine was investigated, and the feasibility was demonstrated through implementation of a smaller scale prototype system. A local leather processing factory, the Hang Yip Leather Factory was consulted to determine the requirements of an automated spraying machine to performing one of the spraying process, lacquer spraying. The expectation of the factory is a compact machine size adaptable to the factory environment and space available, good spraying quality, low frequency of maintenance and cleaning, capability of performing automatic spraying as well as manual controlled spraying. An additional constraint is the need for spark proof operation to minimize chance of explosions.

The leather spray system developed consists of five major sub-systems:

1. leather transport system
2. spraying system
3. environmental isolation system,
4. control system
5. structure/frameworks

The system and design and implementation are described in detail in the following sections. The design of such production equipment requires a integrated mechatronic design approach [1, 2, 3, 4, 5].

2. SYSTEM OVERVIEW

The major components and subsystems are shown schematically in Figure 1. Aluminum profiles comprise the major structural elements and framework. A table transports leathers into an enclosed spraying chamber. During the transport, the size and shape of the leather is sensed with optical sensors.

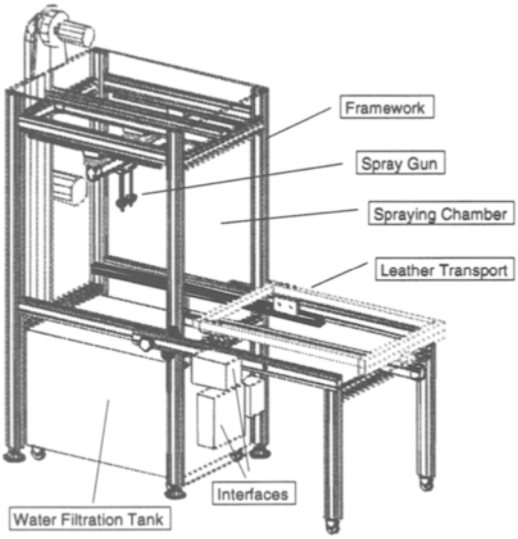


Figure 1. Schematic of Leather Spray System Design

A spray gun applies the coating in a raster-like motion provided by X-Y motion pneumatic cylinders. During the spraying of the leather, fresh air is drawn in to push away the accumulated fumes. The air is drawn through a waterfall filtration system and exhausted externally. Various interfaces are gathered together into a control/interface box. A PC with a digital I/O card is used as a controller. The user interface consists of a control panel and PC display of system status. Two modes of operation are provided, automatic spraying and manual spraying.

3. DESIGN CONSIDERATIONS

The design process identified a number of system requirements:

1. A compact system capable of spraying leathers of irregular shapes and sizes of up to 1.3 meters by 2 meters in size.
2. Wastage of paints/coatings during the spraying process should be minimized to conserve paints and keep operating costs low.
3. Automatic and manual operation should be possible.
4. Quality of coatings applied automatically should be comparable to quality of human operator applied coatings.
5. The cycle time for automatically spraying a leather should not be greater that cycle time of a skilled human operator.
6. The spray gun nozzle distance to the leather should be operator adjustable.
7. The system should be easily controlled by an operator after a short training/familiarization period.
8. A high level of skill should not be required of a human operator.
9. The system should not produce sparks which might ignite the solvents in the coatings.
10. The operator should be protected from the effects of lacquer solvents and solid suspended particulates generated by spraying.
11. The cost of the system should be low.
12. The system should be reliable and have low maintenance requirements.

Due to space restrictions of the available laboratory facilities, a prototype was implemented at a scale of 40% of the linear scale of the actual leather size capacity required. A number of different design concepts were considered. One major consideration was orientation of the leather for spraying. In the manual spraying process, the leathers are mounted in a frame vertically, or at a slight incline to the vertical. The is to accommodate the physical reach of a human carrying a spray gun. A horizontal placement of the leathers for the automatic system developed was selected to eliminate the needs to mount the leather in the frame with clips. Gravity and friction are sufficient to keep the leather flat and stationary relative to the transport table. The disadvantage of horizontal leather placement is the potentially larger footprint of the system. The disadvantage of the larger footprint was minimized by incorporating the water filtration system below the spraying chamber, and making the leather transport table retractable.

Another consideration is the corrosive effect of the paint and coatings solvents. In particular, mate-

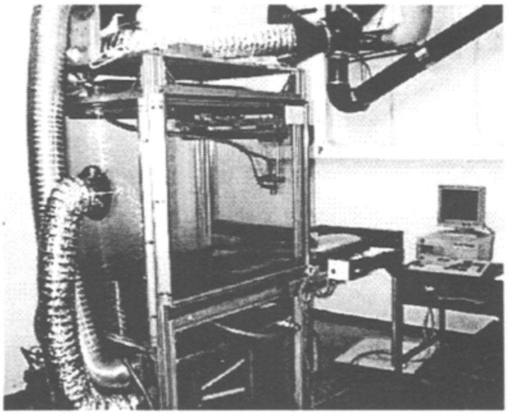


Figure 2. Prototype Leather Spray System.

rials which are in prolonged contact with the fumes need to be selected to be solvent resistant. The prototype developed is shown in Figure 2.

4. FRAMEWORK

The supporting structure for the machine can be divided into the front and back sections. The front section is used to support the leather transportation system and the back section is used to support the spraying system, environmental isolation system and pneumatic components box. In addition, the front section can be retracted to reduce the machine size during the storage and transportation. For the prototype system, aluminum profile was employed to facilitate development.

5. LEATHER TRANSPORT SYSTEM

The leather transportation system consists of a pneumatic cylinder, transport table and tracks. The table moves the leather into and out of the spraying chamber. Figure 3 shows a view of the table with a leather and a design drawing showing pneumatic cylinder and rollers.

A table frame of aluminum profile carries a series of braided steel cables to provide a support surface. The cables provide strength while not accumulating an excessive amount of paint or lacquer. The cables run only lengthwise between the optical sensors to avoid false readings of the leather shape. Frictional forces are adequate for preventing slippage during motion. If additional stability is required clips may be used to fasten the leather to the steel cables. The table frame is supported by two sets of rollers and the pneumatic cylinder. A release mechanism allows the transport table and the supporting slides to be retracted under the spraying chamber when not in use.

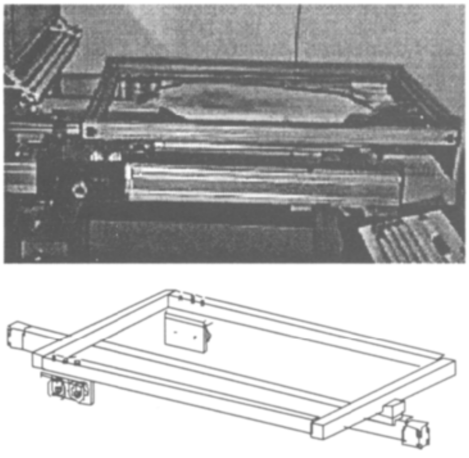


Figure 3. Leather Transport Table.

6. SPRAYING SYSTEM

The spraying system provides X-Y motion of the spray gun. The spraying system consists of the guide system, X and Y axes pneumatic cylinders, spray gun and its mountings. Figure 4 shows the pneumatic cylinders and guides for travel in the longer X direction and the spray gun mounted under the Y axis pneumatic cylinder.

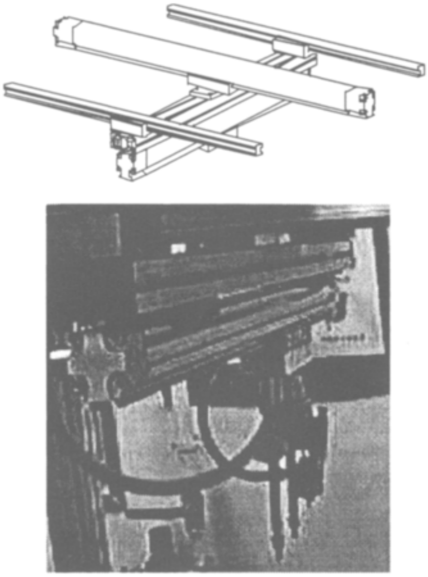


Figure 4. Pneumatic Cylinders for Spray Gun X-Y Motion.

7 ENVIRONMENTAL ISOLATION SYSTEM

The environmental isolation system filters fumes (which may consist of solid particulates and solvents) produced by spraying, and vents the filtered gases from the spraying chamber to the outside area. The environment isolation system consists of a ventilation and filtration system, and sealing. Ventilation and filtration is provided by a water tank, water pump, air blower and accessories such as ducts, flanges water pipes and valves. The sealing system includes an air curtain, rubber seals and tempered glass (for safety) windows enclosing the spraying chamber. An air curtain prevents fumes from venting out of the spraying chamber when the table is outside. Rubber seals prevent fumes from escaping when the table is inside the spraying chamber.

Air is drawn out of the spraying chamber

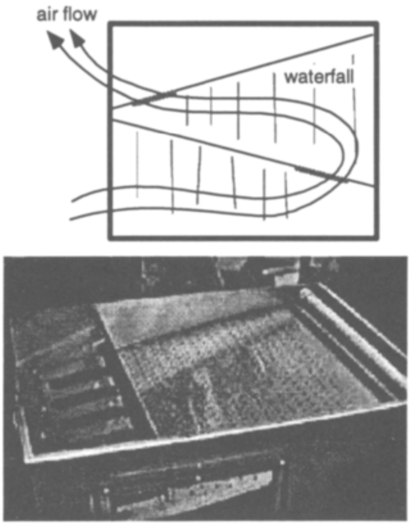


Figure 5. Waterfall Filtration Tank.

during spraying and through a waterfall filtration tank. As the air passes through the waterfall, both solid particulate and solvents are removed. The air is then vented externally. The water is recirculated and can be drained by an operator when required. An acrylic door on the filtration tank allows an operator to judge the need for a water change by viewing the turbidity. The waterfall is created with two inclined plates within the filtration tank. A regular pattern of holes drilled in the plates creates a rainfall effect. The rainfall and air circulation is illustrated in Figure 5. The tank is constructed using welded stainless steel. A view of the first inclined plane is shown in Figure 5 with the top cover of the tank removed.

8 CONTROL SYSTEM

The control system includes a central controller and subsystems for spraying, leather transport table, leather shape/pattern sensing, environmental isolation, interfacing, sensors and control software.

8.1 Spraying Control

The spraying control system uses proportional pneumatic valves to control the speed of the pneumatic cylinders carrying the spray gun. Inputs from 0 to 10 volts control both the speed and direction of motion of the pneumatic cylinders. The centre point of 5 volts results in no motion. On either side of the 5 volt level, speed from minimum to maximum are provided in two different directions. Rather than implementing continuous speed control, five discrete speeds are provided by voltage regulators: high and low in both directions as well as for stopping. A small number of discrete speeds is adequate for the spraying process and helps to conserve digital I/O lines.

The position feedback of the X-Y position of the pneumatic cylinder is provided by sensing magnets placed on the X-Y slides. A series of reed switches are mounted on the X and the Y axis pneumatic cylinders for position detection. Figure 6 shows the magnet attached to the X direction slide and some of the reed switches for sensing X position which are normally hidden by a rubber seal. The spraying gun is actuated by a solenoid.

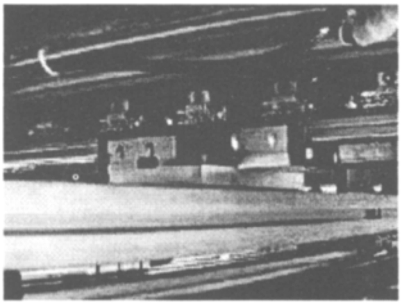


Figure 6. Reed Switches for Spray Gun Positioning

8.2 Transport Table Control

The leather transport table is actuated by a pneumatic cylinder. Unlike the spray gun, variable speed is not required. Two reed switch sensors mounted on the both ends of pneumatic cylinder sense magnets attached to the transport table frame to allow sensing of the table position, inside or outside of the spraying chamber.

8.3 Leather Shape Sensing

The shape of the leather is sensed with a row of optical sensors as the leather is transported into the spraying chamber. The sensors are placed at 10 cm centres spacing. Due to the size of the spray pattern footprint of the spray gun nozzle, high resolution shape sensing is not required. The sensors are read at regular intervals of the table motion into the spraying chamber to generate a grid pattern representing the leather shape. Figure 7 shows the row of optical sensors for detecting the leather.

In order to sense the motion of the table, magnets attached to the transport table frame are sensed with a reed switch. This allows slight variations in speed to be compensated for. As a magnet passes the reed switch during motion into the spraying chamber, a reading of the optical sensors is triggered. Figure 8 shows the magnets and reed switch sensor.

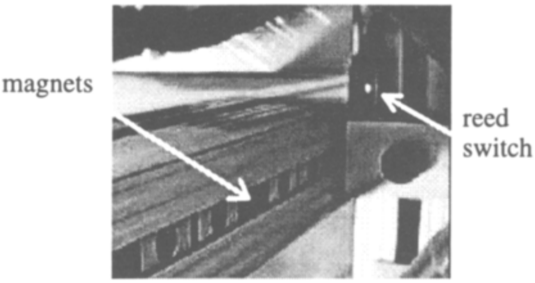


Figure 8. Table Position Sensing to Trigger Leather Sensing

The layout of the leather pattern sensors is illustrated in Figure 9.

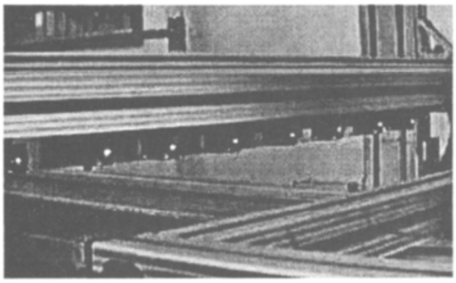


Figure 7. Optical Sensors for Detecting Leather

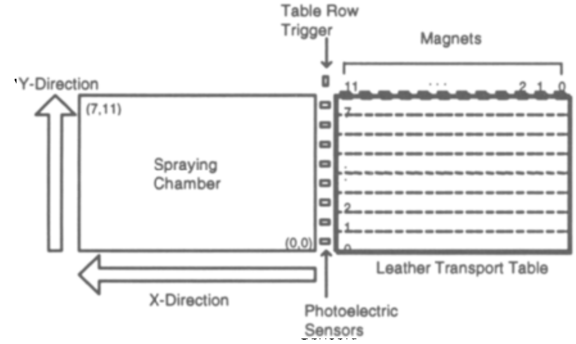


Figure 9. Sensing of Leather Pattern.

8.4 Ventilation and Filtration Control

Ventilation control circuitry consists of two parts. The first part is an on/off control of environmental isolation system's air blower and water pump. The second part is the speed control of air blower. During the ready state, the air blower will operate at low speed to remove residual fumes but when the spraying operation starts, the air blower will operate at high speed. A frequency controller allows digital control of the blower speed. A power saving routine in the control program turns off the environmental isolation system to conserve energy if there is a prolonged idle period.

8.5 Central Controller

A PC with a digital I/O card is used as the central controller. A standard PC is used as a controller for convenience of program development. The current controller is an Intel 80386 based system with 1 megabyte of memory running the DOS operating system. An Intel 80286 based PC was used and found to have adequate speed. A dual 8255 I/O card provides a total of 48 digital input/output lines. The controller for the prototype system requires 21 input lines and 24 output lines. The number of spare input lines is adequate for supporting the increase in inputs required for the full scale model. Encoders and decoders are used to reduce the number of I/O lined needed. In addition to the motion control and leather pattern sensing, the controller also implements the user interface through a control panel and monitor display.

The PC and digital I/O card combination controller is very economical (the PC obtained for the controller was a surplus system). Migration to an Intel 80386 embedded controller at a later stage would be a relatively simple task. A control program is written in the C programming language. A state machine controller is implemented using polling loops. The operational speed of the system did not require interrupt based programming.

The control program comprises of two major polling loops for automatic and manual control. At power on, an initialization routine moves the spray gun to a known position and the table to the exterior position.

The automatic spraying mode causes the spray gun to trace over the leather in raster like motion while applying the coating. The spray gun is actuated only where the leather has been detected. To give a good edge finish, the motion extends one grid point beyond the last and first positions of detected

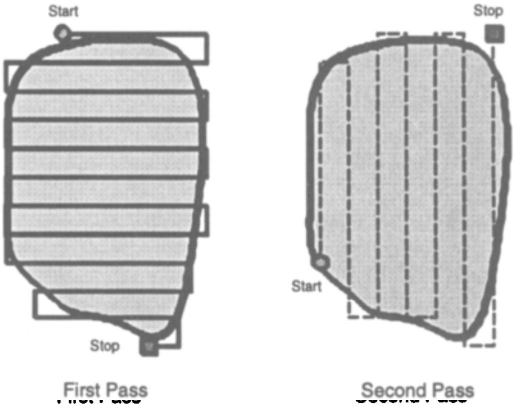


Figure 10. Two Pass Spraying of Leather.

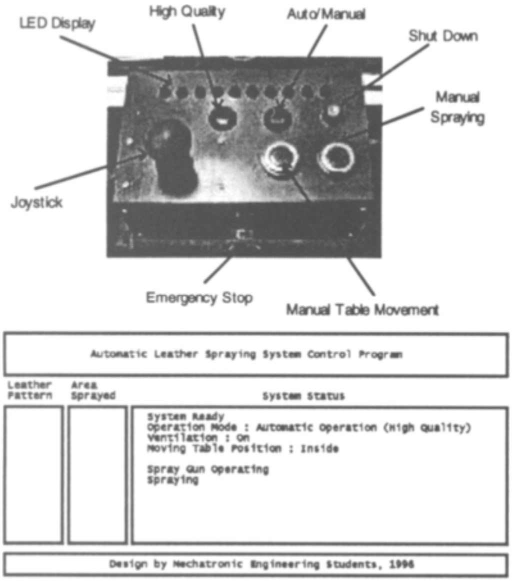
leather on a scan line. No spraying occurs in interior holes and during the step to the next scan line. A high quality spraying mode can be selected which performs a second pass of the spraying gun in the direction perpendicular to the first, illustrated in Figure 10.

8.6 User Interface

The user interface is shown in Figure 11 consists of an operator's control panel and computer monitor display. The operator panel includes: joystick for manual control of spray gun movement, manual table movement button, manual spraying button, auto/manual operation mode switch, dual automatic operation start button (the pair of start buttons requires two handed operation for safety), automatic high quality spraying switch, shut down button, emergency stop button and a set of LEDs for error/system status display. A computer monitor display provides additional information on system status and realtime graphical display of automatic spraying operation progress..

Two different operating modes (automatic and manual) can be selected at any time during usage. A set of LEDs will display the current error/system status. For emergency events or any incorrect operation, the system can be halted immediately by pressing an emergency stop button which cuts off central supply of compressed air to the pneumatics.

Manual operation mode allows the operator to control the spraying, spray gun motion and transport table motion using the manual spraying button joystick and manual table movement button respectively.



```
Automatic Leather Spraying System Control Program

System Ready
Operation Mode : Automatic Operation (High Quality)
Ventilation : On
Moving Table Position : Inside

Spray Gun Operating
Spraying

Design by Mechatronic Engineering Students, 1996
```

Figure 11. User Interface.

9. ACKNOWLEDGMENTS

The following BEng in Mechatronic Engineering students have contributed to the design and implementation of the prototype system: Chan Tak Fai, Chan Wai Man, Cheung Pok Yeung, Lam Wing Cheung, Chan Kam Lun, Cheung Kwok Wai, Li Sze Lui, Poon Kok Way, Li Chi Kong.

10. CONCLUSIONS

A case study of a successful design of a mechatronic production equipment has been described. The major subsystems designed and the control integration are described briefly. The final prototype was capable of spraying a leather in approximately 1 minute, comparable to speed of a human operator in manual spraying. Testing of the environmental isolation system showed that fumes are successfully gathered and vented externally after filtration. Based upon the successful implementation of the prototype, it can be concluded that success is indicated in the feasibility of the system design concepts for application to a full scale system. The success of the design and implementation required a careful system design and analysis. A number of peripheral problems have not been described including electrical noise suppression, cable management, etc.

REFERENCES

[1] Hewit, James R. King, Tim G. "Mechatronics design for product enhancement", IEEE/ASME Transactions on Mechatronics. v 1 n 2 Jun 1996. p 111-119.

[2] Taylor, P M., Gunner, M B., Smith, D R., Adams, E J., Song, X K., Hu, X X., Taylor, G E. "Mechatronic design of an automated garment production line", IEE Colloquium (Digest). n 191 1994. p 3/1-4.

[3] Valasek, M. "Inventive engineering as basic of design methodology for mechatronics", Proc Jt Hung Br Int Mechatronics Conf 1994. p 563-566.

[4] Walsh, Ronald A. Electromechanical design handbook. McGraw-Hill, 1995.

[5] Pugh, Stuart. Total design : integrated methods for successful product engineering. Addison-Wesley, 1991.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Mechatronic System for the Improvement of Surface Form in Planed and Moulded Timber Components

N Brown B.Eng M.Sc,
R M Parkin B.Sc Ph.D FRSA C.Eng MIEEE FIEE F.I.Mech E
Mechatronics Research Group, Systems Engineering Research Centre, Faculty of Engineering,
University of Loughborough LE11 3TU, U.K.

Planing and moulding operations carried out within the woodworking industry make extensive use of rotary machining. Cutter marks are produced on the timber surface which are generally accepted as unavoidable. When a high quality finish is required however, a further machining operation, such as sanding, is often required to remove cutter marks. It has been proposed that surface form of the timber surface may be improved by oscillation of the cutterblock in order to obviate the need for further machining. This paper describes the development of a system for cutterblock oscillation for this purpose. Results are promising, and show clearly the effectiveness of this method for improving surface quality.

NOMENCLATURE

A	Hydraulic ram area (m^2)
B	Bulk modulus ($N.m^{-2}$)
fd	Feedrate ($m.min^{-2}$)
$l(t)_x$	Instantaneous cutter tip locus (x axis)
$l(t)_y$	Instantaneous cutter tip locus (y axis)
k_A	Drive amplifier constant ($mA.V^{-1}$)
k_{LE}	Leakage coefficient ($cm^3/N.m^{-2}$)
k_s	Spool displacement gain ($mm.mA^{-1}$)
M	Ram, slideway, cutterblock mass (kg)
p	Pressure ($N.m^{-2}$)
μ	Coefficient of viscous friction
r_{ct}	Radius of cutting circle (m)
s	Laplace operator
v_m	Valve actuator time constant
V	Volume of oil column (m^3)
V_v	Valve voltage (V)
ω_v	Valve natural frequency (rad/s)
x	Hydraulic ram displacement (m)
X_s	Valve spool displacement (m)
ζ_v	Valve spool damping

1. INTRODUCTION

The rotary wood machining process is similar in nature to the upcut milling of metals. There are however, marked differences between the two applications, the primary one being cutting speed, which lies in the range $30m.sec^{-1}$ - $80m.sec^{-1}$ compared with typically $0.5m.sec^{-1}$ - $1.5m.sec^{-1}$ for the milling of metals [1]. The feed speed in wood machining is correspondingly high ranging from $5m.min^{-1}$ - $100m.min^{-1}$. Figure 1 shows the general layout of a 2 head (top and bottom) planing / moulding machine.

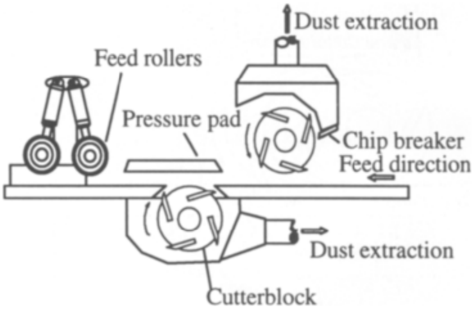


Figure 1. General layout of planing/ moulding machine

Planed and moulded surfaces appear, when viewed closely, as a series of waves whose peaks are perpendicular to the passage of the product through the machine (figure 2). Machining errors may exist which affect surface finish, these being surface waviness and error of form. Waviness is the component of texture upon which the undulating knife traces (roughness) are superimposed, and can result from cutterhead imbalance, the flexure of the machine frame, and machine vibrations.

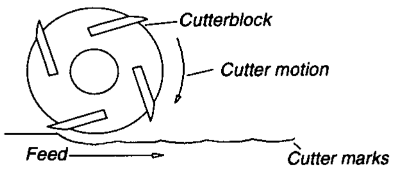

Cutterblock
Cutter motion
Feed
Cutter marks

Figure 2. Effect of rotary machining on surface

Fixed knife planing has been shown to produce an excellent surface finish whereby none of these defects are present. The high feed forces and feed methods used however may mark the product, and only a small amount of material may be removed using a fixed knife, around 0.2mm [1], hence fixed knife machining is not suitable for moulding.

2. NOVEL CUTTERBLOCK MOTION

It was intended that the effectiveness of a modified form of cutterblock motion could be verified. Figure 3 shows how a cutterblock with two degrees of freedom may reduce cutter wave height. The cutter advances producing a clean cut (figure 3 - displacements are exaggerated for clarity).

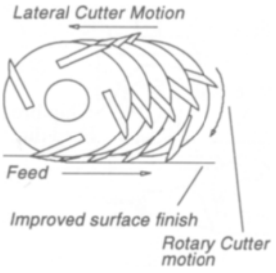

Lateral Cutter Motion
Feed
Improved surface finish
Rotary Cutter motion

Figure 3. Improved surface due to horizontal cutter motion

2.1. Actuator performance requirements

The stroke of the actuator system is determined by the cutter wave pitch .For average quality material cutter wave pitch figures in excess of 2.5mm are commonplace, though for high quality products, wave pitch should be 1.0mm or less [1].

High frequencies of actuation are required, for example a 4 knife cutter rotating at 6000 rev/min requires 4 oscillations of the cutter head per revolution, or 24000 oscillations/min.

3. SERVOACTUATOR SIMULATION

Hydraulics were seen as the best choice of actuator, whereby hydraulic cylinders would be controlled by valves capable of operating at high frequencies, such as servo valves. The complete hydraulic circuit is simulated, including the hydraulic source, servo valve and hydraulic load (Figure 4). Effects peculiar to a hydraulic system are included such as fluid bulk modulus, oil column stiffness, and pressure dependent friction, which are needed to produce a comprehensive model. [8,9,10].

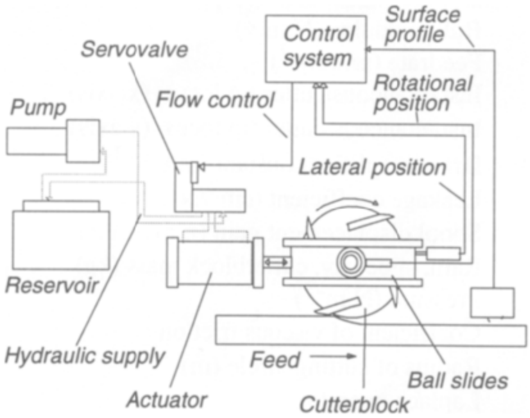

Surface profile
Control system
Servovalve
Rotational position
Pump
Flow control
Lateral position
Reservoir
Hydraulic supply
Feed
Ball slides
Actuator
Cutterblock

Figure 4. Hydraulic system schematic

3.1. Actuator Simulation

Since manufactures transfer functions are available (e.g. [4]), transfer functions of the valves are utilised, determined by manufacturer's practical testing of the servo valves. (eqn 1). The transfer function shown in (eqn 2) is used to model the hydraulic ram and slideway.

$$\frac{Q(s)}{V_v(s)} = \frac{k_a}{1+v_m \cdot s} \cdot \frac{k_s}{\frac{s^2}{w_v^2} + \frac{2z_v}{w_v} \cdot s + 1} \qquad (1)$$

$$\frac{X(s)}{Q(s)} = \frac{1}{\frac{MV}{BA}s^3 + \left(\frac{k_{LE}M}{A} + \frac{\mu V}{BA}\right)s^2 + \left(\frac{k_{LE}\mu}{A} + A\right)s} \qquad (2)$$

In order to derive the ram - slideway transfer function, a single acting hydraulic ram is considered as shown in figure 5. The piston cross sectional area is taken as A, the input fluid flow to the ram is given by q(t), p(t) is the fluid pressure and the ram position is given by x(t).

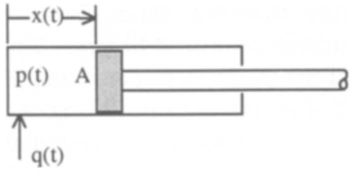


Figure 5. Hydraulic ram

In order to provide a comprehensive model, it is necessary to consider the mass moved, (i.e. ram, slideway and cutterblock), any leakage which may exist between ram and cylinder wall, effects of viscous friction, and the compressibility of the fluid, i.e. the fluid bulk modulus. The governing equation for a simple ram is given by:

$$q(t) = A\frac{dx(t)}{dt} \qquad (3)$$

Taking into account leakage, where a leakage flow of $k_{LE}p(t)$ exists, the flow continuity equation then becomes:

$$A\frac{dx(t)}{dt} + k_{LE}p(t) \qquad (4)$$

Calculation of $p(t)$ due to inertia is achievable through applying Newton's Second Law, i.e. *Force = Mass x Acceleration.*

$$p(t)A = M\frac{d^2x(t)}{dt^2} \qquad (5)$$

The ram will be subject to a viscous load, whereby the coefficient of viscous friction μ affects the force available from the ram. The pressure $p_f(t)$ required to overcome this frictional loss is therefore:

$$p_f(t) = \mu\frac{dx(t)}{dt} \qquad (6)$$

It is then necessary to consider the effects of fluid compressibility. Bulk modulus may be defined as change in fluid pressure divided by change in volume per unit volume, i.e.

$$B = \frac{\Delta p}{\Delta v} \quad (7) \qquad \therefore \frac{dv}{dt} = \frac{V}{B}\frac{dp}{dt} \quad (8)$$

The complete flow continuity equation is therefore:

$$q_{in} = q_{vel} + q_{leakage} + q_{compressibility} \qquad (9)$$

$$q(t) = A\frac{dx(t)}{dt} + k_{LE}p(t) + \left(\frac{V}{B}.\frac{dp(t)}{dt}\right) \qquad (10)$$

Viscous friction and inertia are used to calculate the pressure required to move the ram:

$$p(t)A - \mu\frac{dx(t)}{dt} = M\frac{d^2x(t)}{dt^2} \qquad (11)$$

$$\therefore q(t) = A\frac{dx(t)}{dt} + k_{LE}\left(\frac{M}{A}\frac{d^2x(t)}{dt^2} + \frac{\mu}{A}\frac{dx(t)}{dt}\right) + \frac{V}{B}\left(\frac{M}{A}\frac{d^3x(t)}{dt^3} + \frac{\mu}{A}\frac{d^2x(t)}{dt^2}\right) + \left(\frac{k_{LE}\mu}{A} + A\right)\frac{dx(t)}{dt} \qquad (12)$$

$$\therefore q(t) = \frac{VM}{BA}\frac{d^3x(t)}{dt^3} + \left(\frac{k_{LE}M}{A} + \frac{\mu V}{BA}\right)\frac{d^2x(t)}{dt^2} \qquad (13)$$

Taking Laplace transforms for zero initial conditions gives the ram - slideway transfer function:

$$\frac{X(s)}{Q(s)} = \frac{1}{\frac{MV}{BA}s^3 + \left(\frac{k_{LE}M}{A} + \frac{\mu V}{BA}\right)s^2 + \left(\frac{k_{LE}\mu}{A} + A\right)s} \qquad (14)$$

4. SURFACE FINISH SIMULATION

The simulation of the cutting process is carried out using Matlab software. An ideal cutter tip locus is first described using equations (15), and (16).

$$I(t)_x = r_{ct}.\cos f_d + \left(\frac{2\pi n}{60}\right) - r_{ct}\cos(f_d) \qquad (15)$$

$$I(t)_y = r_{ct}.\sin f_d + \left(\frac{2\pi n}{60}\right) - r_{ct}\sin(f_d) \qquad (16)$$

The cutter tip locus is then processed using software outlined in figure 5. This produces a model which takes into account cutterhead imbalance, the effect of a proud cutter knife, and machine vibration, but improves upon previous models [5,6,7] in that the effect of external vertical or horizontal waveforms may be predicted.

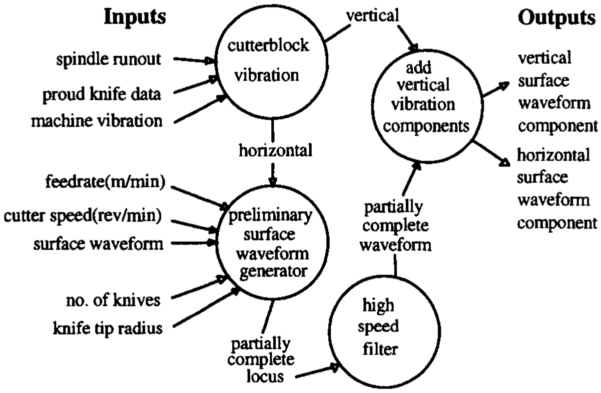

Inputs
spindle runout
proud knife data
machine vibration
cutterblock vibration
vertical
horizontal
feedrate(m/min)
cutter speed(rev/min)
surface waveform
no. of knives
knife tip radius
preliminary surface waveform generator
partially complete locus
add vertical vibration components
partially complete waveform
high speed filter
Outputs
vertical surface waveform component
horizontal surface waveform component

Figure 5. surface simulation software data flows

The high speed filter was designed to give the software developed in Matlab equivalent performance to software previously developed in C. The cutter tip locus is scanned to produce the waveform as shown in figure 6.

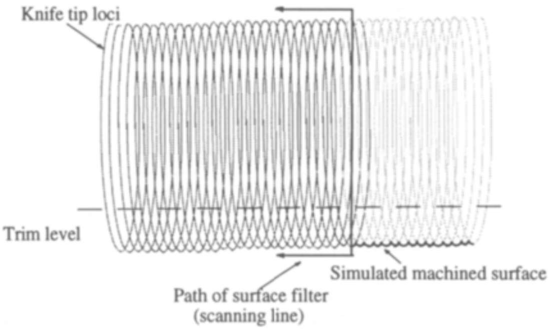

Knife tip loci
Trim level
Simulated machined surface
Path of surface filter
(scanning line)

Figure 6. Extraction of surface data.

The trim level shown in the figure is reached through using a sorting algorithm which forms a standard part of the matlab command library. The vertical components of the entire locus are sorted in ascending order, with a separate array storing the locations of the corresponding horizontal components. The order of these components is not important since the sorting algorithm will still scan through the vertical components within a fragmented equivalent of a 2 dimensional variable space (i.e. not every location in the variable space is filled). Figure 7 shows the effect of using a fixed sampling period to produce a waveform with regular horizontal steps. In practice such a method generates a particularly messy output waveform, often with spurious samples having vertical components from the upper curves of the cutter tip locus. The use of a large number of loci points appears to overcome this problem in part, but the appearance of spurious samples is of course not completely overcome, and the large arrays used to hold the cutter loci are too unwieldy for computation. An array of around one Mb in length for the cutter loci was found to be capable of producing a 1000 point regularly spaced horizontal sample with around 50 spurious samples. The time spent processing this data would be inordinate, unless manual cleaning up of the results was deemed acceptable.

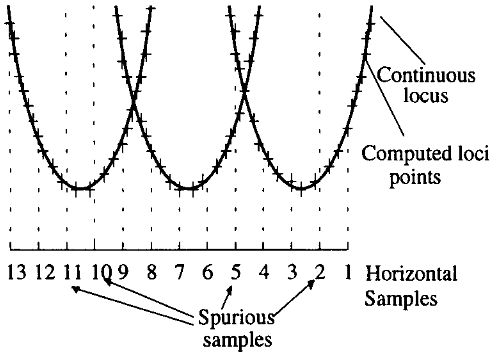


Figure 7. Spurious data produced by unevenly spaced horizontal loci.

The main control loop of the filtering software initially specifies which output sample is to be taken. A scan is then made of which loci points are above the sample point. A small tolerance is built in such that a specified number of sample points which come nearest to the vertical sampling line are taken only. A high speed bubble sort algorithm is used to sort the surrounding area in order of distance from the sample line. A fixed number of the lowest sample points in the list are taken.

This list is then sorted using the same algorithm in order of vertical height. Ideally, the first sample point in the list would always be the lowest loci component, and therefore the appropriate point to be stored in the horizontal sample array. The same loci sample may be shared by one or more inputs to the horizontal sample, so it is necessary to successively pull in the spread of loci samples until it is clear that the lowest loci sample nearest to the sample line has been found.

For example, referring to figure 8, it is clear that sample point vl is the closest, and therefore most appropriate sample to take in terms of accuracy at the vertical line marked by sample 5. Vertically spurious samples are easily produced by existing software: a simpler algorithm may pick up the loci point closest to the sample line, which in this case would be v6, which would clearly be a spurious sample, giving an incorrect picture of the surface waveform. Horizontally spurious samples must be avoided by taking successively smaller pictures of the surrounding area to the sample line, e.g. in the first wave of sampling, in general all v samples as well as wl -7, and in particular samples w8 - 10 would be produced as candidates for the lowest sample point. The code contained in the while loop in the filtering software is able to home in on the sample point V1 in order to alleviate this. The final iteration may produce samples vl, v2, v6 and v7. It is then very straightforward to pick the lowest of these samples.

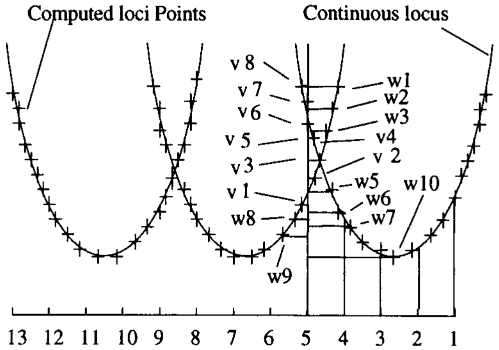


Figure 8. Sampling procedure

Code then re sorts the nearest loci points to the sample line and checks the lowest value against that found in the preceding loop. When the value of the vertical loci component of the lowest perceived loci sample starts to rise in the following loop, it is clear that a minimum vertical loci point has been found with no 'smearing' of the results from one fixed horizontal point to the next.

Consequently the sampling method is completely robust, and changes in the spacing of the loci points will not affect the performance of the sampling code, which adjusts automatically.

5. RESULTS

5.1. Hydraulic system

The plot shown in figure 9 shows the simulated response of the hydraulic servo. As can be seen, the second order transfer model provides a reasonable approximation to the true servo valve response for the frequency range of interest, i.e. 50 - 150 Hz.

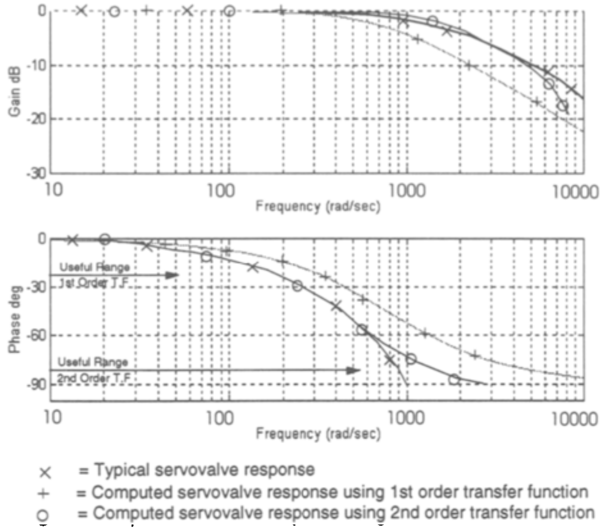


Figure 9. Servo valve response bode plot

Figure 10 shows the predicted output of the hydraulic servo using a square wave as the command signal. The use of the square wave introduces odd harmonics which can be seen to have interesting second and third order effects upon the system. Whilst acceleration is slightly improved when compared to a sine-driven simulation, the use of a clipped sine wave is adopted for the controller since this interfaces readily with transducers which detect cutter position in the proposed final design.

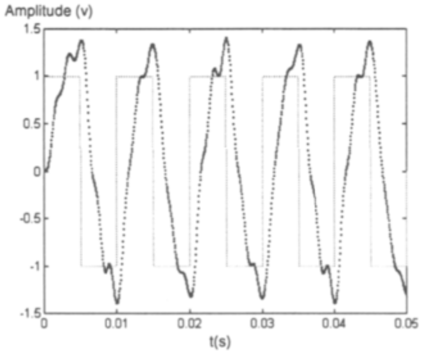


Figure 10. Servo response (square)

Further tests have revealed that the closed loop system is stable for gain values which place the magnitude of the servo response within the required limits for cutterblock oscillation. With the correct design of controller in place, the servo output should be stable and drift should be negligible.

5.2. Surface Finish

The simulated conventionally produced surface waveform may be seen in figure 11. An improvement in surface finish may be seen in figures 11 - 14 The actuation waveform comes from a clipped sine wave which is generated from the vertical component of the cutter tip locus.

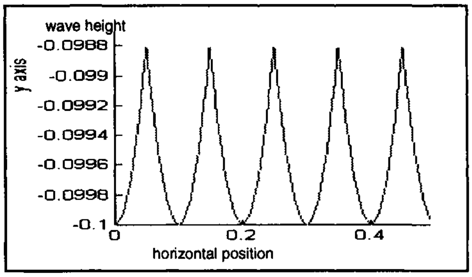


Figure 11. Simulated surface wave for 0% oscillation

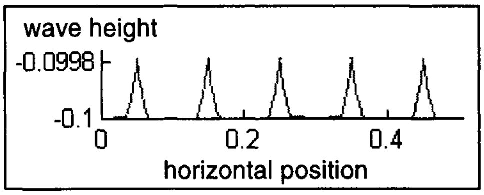


Figure 12. Simulated Surface wave for 40% oscillation

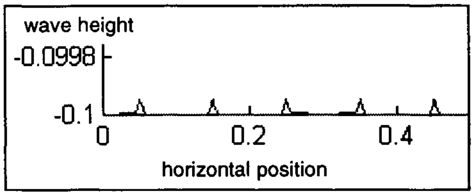


Figure 13 Simulated Surface wave for 60% oscillation

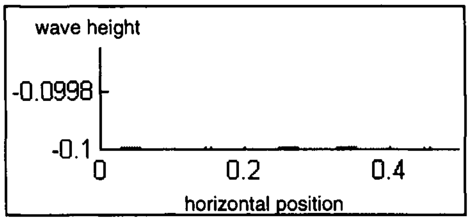


Figure 14. Simulated Surface wave for 80% oscillation

Experimentally obtained results using a mechanical simulator for cutterblock oscillation may be seen in figure 15. A sine wave of amplitude equal to 50% of the feed per revolution has been used to achieve a reduction in wave height of 50%.

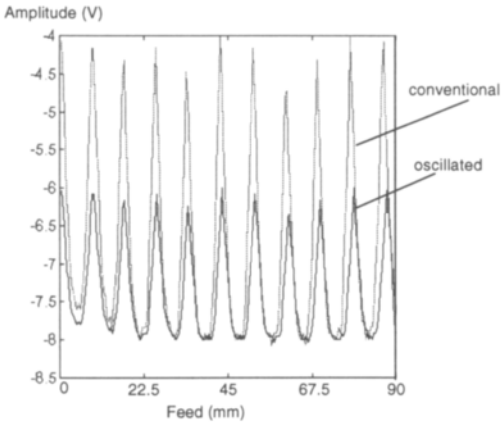


Figure 15. Experimentally obtained surfaces

6. CONCLUSIONS

A surface simulation algorithm has been developed which may be used to model the effect of cutterblock oscillation, and unintended vibration due to cutterblock imbalance. The algorithm is flexible enough to include the effects of the horizontal position data obtained from the ram - slideway simulation upon the surface finish.

The usefulness of simulating the system has been examined and verified, paving the way for a physical prototype machine, computer modelling having highlighted certain aspects of the hydraulic system to which particular attention must be paid as the design process continues. It has been decided that a hydraulic system will provide the best means of cutter block actuation, and the suitability of hydraulics for such a high - speed system has been proven.

The simulation of the cutting process suggests strongly that the modified machining method can be made to work.

Accurate sequencing of the actuation wave with cutter rotational position is necessary to ensure improvements in surface finish with further refinements of the system. Future work includes the integration of both cutter and actuator simulations, and the use of an intelligent controller to adjust for drift in the system, thereby producing a consistent surface form of high quality.

REFERENCES:

1. M.R Jackson, *Some effects of machine characteristics on the surface quality of planed and spindle moulded wooden products*. Ph. D. Thesis, Leicester Polytechnic and Wadkin Plc. June 1986.

2. P G J Leaney, *The Modelling and Computer Aided design of Hydraulic Servosystems*, Ph.D Thesis, Loughborough University of Technology, 1986

3. T.J Viersma, *Analysis, Synthesis and Design of Hydraulic Servosystems and Pipelines.* , Elsevier, 1980.

4. W. J. Thayler, *Transfer Functions for Moog Servovalves*, January 1963. Moog Technical Bulletin 103. Moog Inc. Controls Division, East Aurora, N.Y. 14052, U.S.A.

5. K.M Maycock. *The Assessment of Surface Quality in Planed and Spindle Moulded Products.* Ph.D. Thesis, DeMontfort University, 1993.

6. F. Ismail, M. A. Elbestawi, *Generation of Milled Surfaces Including Tool Dynamics and Wear.* R. Du, R. Urbasik. Journal of Engineering for Industry, Vol. 115, August 1993.

7. S N Mezzelkote, AR Thangaraj, *Enhanced Surface Texture Model For End Milling.* Winter annual general meeting of the American Society of Mechanical Engineers, Anaheim, CA, USA, 08 - 13/11/92.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Application of Fuzzy PI control to improve the positioning accuracy of a rotary-linear motor driven by two-dimensional ultrasonic actuators

Masato Kinouchi[a], Iwao Hayashi[b], Nobuyuki Iwatsuki[b], Kouichi Morikawa[b]
Jyunji Shibata[c] and Kohji Suzuki[d]

[a]Graduate school, Shibaura Institute of Technology, 307 Tameihara, Fukasaku, Ohmiya, 330-0003 Japan

[b]Faculty of Engineering, Tokyo Institute of Technology, 2-12-1 Ohokayama, Meguro-ku, Tokyo, 152-8552 Japan

[c]Dept. of Mechanical Engineering, Shibaura Institute of Technology, 307 Tameihara, Fukasaku, Ohmiya, 330-0003 Japan

[d]Mitutoyo Cotporation, 1-20-1 Sakato, Takatsu-ku, Kawasaki, 213-0012 Japan

A rotary-linear motor, which is driven by a pair of two-dimensional ultrasonic actuators, has been proposed to drive a small drawing nozzle to draw various patterns on the inner wall of a small pipe several mm's in diameterused to produce precise linear scales. The positioning accuracy of 7 μm was obtained for longitudinal direction drive by applying PI control. However, the positioning response was not steady because the frictional resistance between the tips of two-dimensional ultrasonic actuators and the motor shaft changed rapidly during experiment. Duzzy PI control has hence been applied instead of PI control, and the gain parameters and membership functions were adjusted and some Fuzzy rules were added to improve the positioning accuracy and the stability of response. As the result, the steady error and its standard deviation were reduced to 3.4 μm and 1.9 μm, respectively, for longitudinal direction drive. Also, Fuzzy PI control was found superior to PI control for the frequency response.

1. INTRODUCTION

There is a demand to draw various patterns on the inner wall of a small pipe several millimeters in inner diameter over a span of from 20mm to 30mm to develop a precise small linear scale. To realize this demand, the authors have proposed a rotary-linear motor[1]×[4], in which the motor shaft is driven by a pair of two-dimensional ultrasonic actuators[5] and then can be rotated and moved straightly, simultaneously or independently. A prototype of the proposed rotary-linear motor was produced and the fundamental positioning characteristics were examined constructing a proportional and integral positioning feedback control system. As the result, a good positioning accuracy, 7μm for longtudinal direction drive and 0.05° for rotational drive, was obtained. The response of the proposed motor was, however, unsteady because the motor shaft was driven by friction due to the two-dimensional ultrasonic actuators, and scattered a little according to driving conditions.

To improve this undsteadiness of response and also to drive the rotary-linear motor more precisely, Fuzzy proportional and integral control has been applied, and the parameters were adjusted and some Fuzzy rules were modified and added. The positioning characteristics were experimentally examined through step response and frequency response. The results obtained are thus reported in this paper.

2. ROTARY-LINEAR MOTOR

Figure 1 shows the schematic diagram of rotary-linear motor, which the authors have proposed. The motor shaft is driven through friction by a pair of two-dimensional ultrasonic actuators, which are composed of an aluminium bar 7mm in diameter and four piezostacks adhesived at 45 degrees equally spaced on the circumference of the shoulder of aluminium bar, respectively. The rotational or linear speed of the motor shaft can be independently controlled by changing the duty ratioof chopper pulse for driving pulse with resonant frequency of the two-dimensional ultrasonic actuators.

The photograph of the rotary-linear motor is shown in Fig.2. The motor has a precise linear encoder and a rotary encoder to detect thrust displacement and rotational displacement, respectively. The two-dimensional ultrasonic actuators are supported by parallel plate spring mechanisms with micrometerhead to adjust preload between shaft and actuators.

Figure 3 shows the previous proportional and integral control (hereafter called PI control) system of the rotary-linear motor. It is a position feedback cotrol system, in which the deviation, e_x, is feeded into the PI controller and the duty ratio, r_x, is feeded out as the operating quantity.

3. FUZZY PI CONTROL

3.1 The system of Fuzzy PI cotrol

Figure 4 shows the Fuzzy PI controller[6] which was inserted instead of the PI controller of the previous control system to eliminate the unsteadiness of response of the rotary-linear motor due to friction drive. In the Fuzzy PI controller, the input variables are the deviation of position, E (scaled value), and the time derivation of E, ΔE; the time derivation of duty ratio, ΔR, which is the output of the Fuzzy controller, is calculated and feeded out through Fuzzy inference. For the Fuzzy inference, Fuzzy rules which had a simplified rear part was used, and equally spaced triangular functions were also adopted as the membership functions for the inputs. Figure 5 shows the Fuzzy rules determined based on the step response of the rotary-linear motor.

3.2 Adjustment of input and output gains

The two input gains, G_1 and G_2, and the output gain, G_3, (see Fig.4) were set at first based on the experimental data preobtained by conventional PI control so that the deviation, E, and the time derivation, ΔE, eill not be saturated, and the optimum values were then experimentally determined. The obtained optimum gaines are summarized in Table 1. The rotary-linear motor was tried to drive with these values.

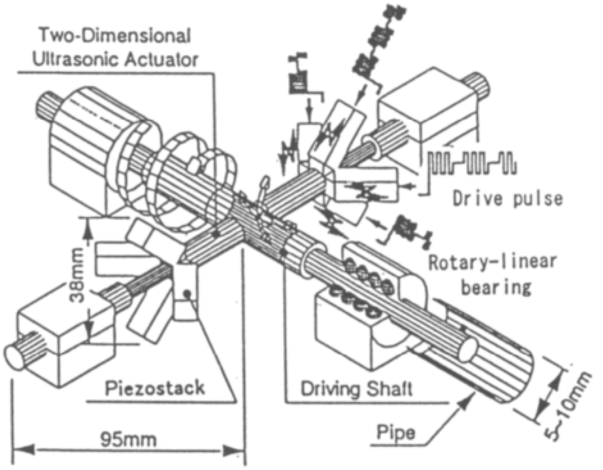


Fig. 1 A schematic diagram of rotary-linear motor

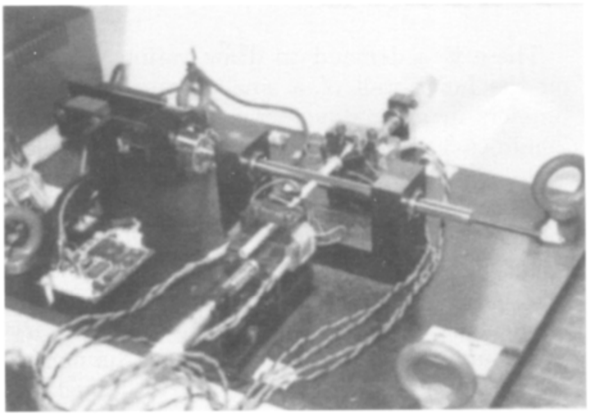


Fig. 2 A prototype of rotary-linear motor

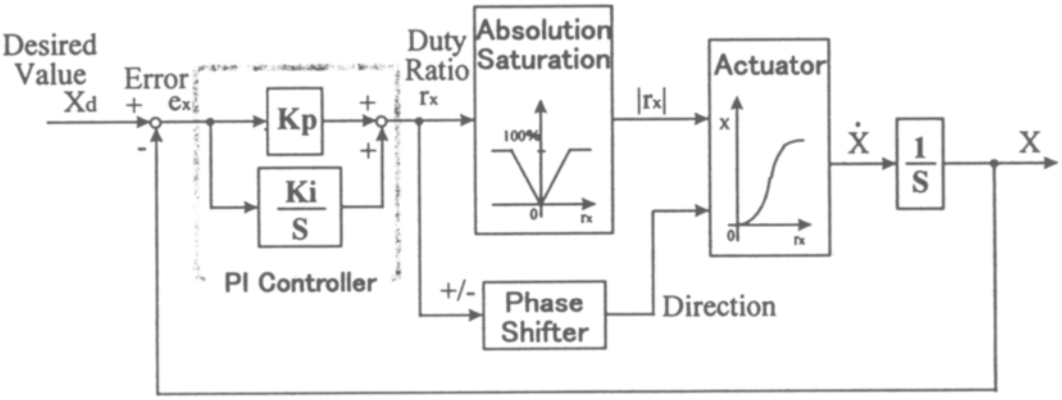

Desired Value
Xd
Error
+ ex
-
Kp
Ki
S
+
+
PI Controller
Duty Ratio
rx
Absolution Saturation
100%
0
rx
|rx|
Actuator
x
0
rx
Ẋ
1
S
X
+/-
Phase Shifter
Direction

Fig. 3 PI control system for rotary-linear motor

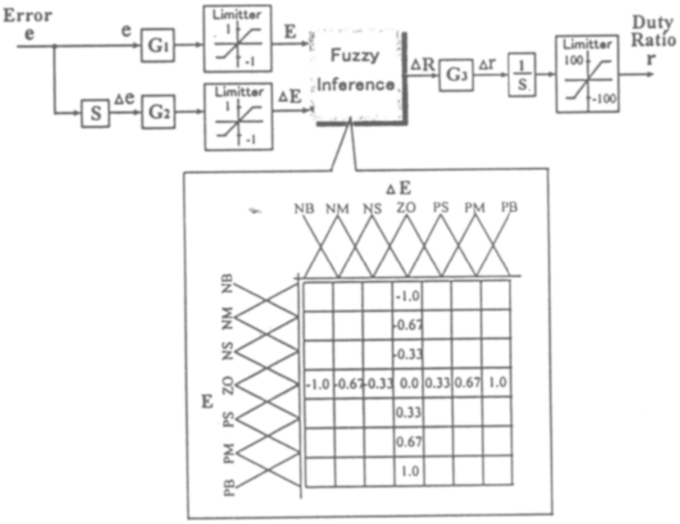

Error
e
e
G1
Limitter
1
-1
E
S
Δe
G2
Limitter
1
-1
ΔE
Fuzzy Inference
ΔR
G3
Δr
1
S
Limitter
100
-100
Duty Ratio
r
ΔE
NB NM NS ZO PS PM PB
E
NB NM NS ZO PS PM PB
-1.0
-0.67
-0.33
-1.0 -0.67 -0.33 0.0 0.33 0.67 1.0
0.33
0.67
1.0

Fig. 4 Fuzzy PI control system for rotary-linear motor

E \ ΔE	NB	NM	NS	ZO	PS	PM	PB
NB				NB			
NM				NM			
NS				NS			
ZO	NB	NM	NS	ZO	PS	PM	PB
PS				PS			
PM				PM			
PB				PB			

Fig. 5 The table of Fuzzy rules

Table 1 Determined optimum input and output gains

	Linear	Rotational
G_1	1/(10mm)	1/(90 °)
G_2	1/(92mm/s)	1/(1406°/s)
G_3	60	

4. DRIVING EXPERIMENTS

4.1 Step response I

The rotary-linear motor was driven in the linear direction using the determined gains. The desired position was set at 10mm and the experiments were repeated five times. Figure 5 shows one example of the five experiments of step response. Table 2 summarizes the postioning error, the standard deviation of the error, and the RMS error after the postion of the motor shaft reached to the desired position within 5% error, which were obtained under the conventional PI control and the proposed Fuzzy PI control. As clearly seen in the table, the positioning error and the standard deviation were improved by appling Fuzzy PI control, but the RMS value was still large. This means that Fuzzy PI control should be improved so that the rotary-linear motor can be controlled more precisely near the desired postion than the present Fuzzy PI control.

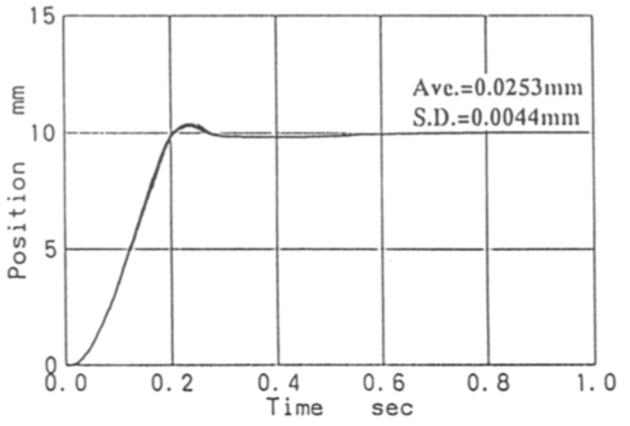


Fig. 6 Examples of step response I for linear drive

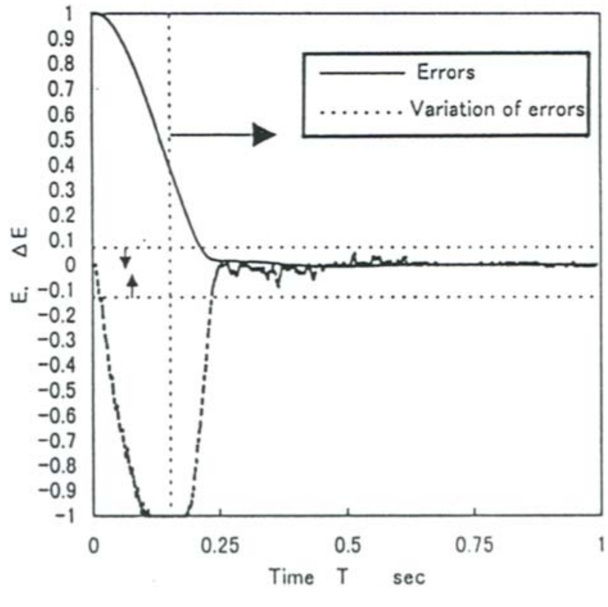


Fig. 7 The time history of errors and its variation

4.2 Modification of membership function

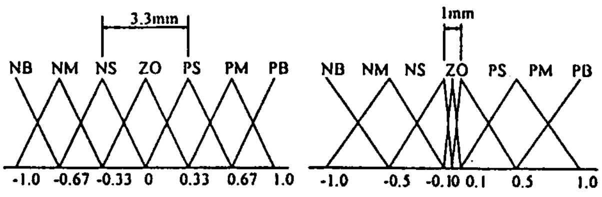
Modification of membership function was tried based on the experimental results mentioned in the above section. The time histories of the two input variables, namely the deviation, E, and the derivation, ΔE, were hence measured. As clearly seen in Fig.7, both of E and ΔE were varying within ±0.1 after they reached to ±0.1. Based on this results, the ZO part of the membership functions was made narrower than in the previous membership functions, namely from 0.33 to 0.1, as shown in Fig.8(a) and 8(b). The range of ZO was accordingly changed from 3.3mm to 1.0mm for the case that the desired position was 10mm. The rotary-linear motor can hence be expected to control more precisely near the desired position.

4.3 Supplment of Fuzzy rules

The two Fuzzy rules, PM and NS as shown in Fig.9, were added to suppress overshoot greatly and to accellerate the starting motion greatly, according to the change of the membership functions. Also, the input gain G_2 was changed to 1/80 only for linear motion.

Table 2 Averaged values of positioning accuracy for linear drive (mm)

	Errors	Standard deviation	R.M.S. values
PI control	0.0731	0.0499	0.091
Fuzzy PI control	0.0253	0.0044	0.121



(a)Before adjustments (b)After adjustments

Fig. 8 Adjustments of membership functions

ΔE

E	NB	NM	NS	ZO	PS	PM	PB
NB				NB	NS		
NM				NM			
NS				NS			PM
ZO	NB	NM	NS	ZO	PS	PM	PB
PS	NS			PS			
PM				PM			
PB			PM	PB			

Fig. 9 The Fuzzy rules after adjustments

4.4 Step response II

The step response of rotary-linear motor was examined again using the modified Fuzzy control. Figure 10 shows one example of the step response experiments, and Table 3 shows the comparison of the positioning errors obtained for the Fuzzy PI control, and the modified Fuzzy PI control. The desired position was 10mm and the experiments were repeated five times as previously. As clearly seen in the table, the RMS value was reduced greatly, and the positioning error and the standard deviation were also improved greatly; In other words, the unsteadiness has been greatly improved by applying Fuzzy PI control and adjusting the parameters.

4.5 Frequency response

The frequency response of the rotary-linear motor was also examined by applying Fuzzy PI control. Figure 11 shows the practical Fuzzy controller for frequency response; Tbale 4 shows the determined input and output gains. The measured bode diagrams of frequency response are shown in Fig.12. As clearly seen in the figure, the frequency response obtained by appling Fuzzy PI control was better than obtained by PI control for both of gain and phase delay in the frequency range from 0.5Hz to 1.0Hz; this means that frequency response has been improved by applying Fuzzy PI control, too.

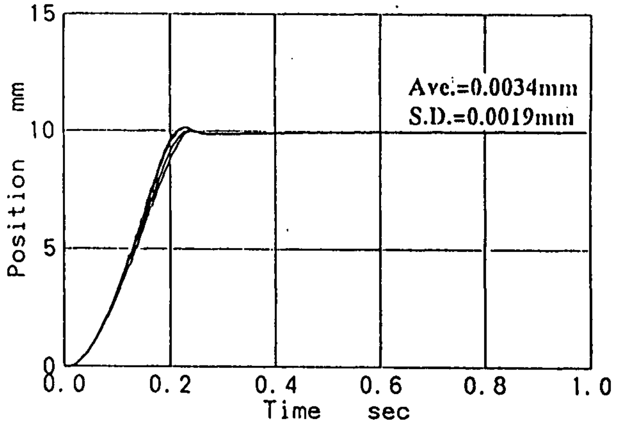


Fig. 10 Examples of step response II for linear drive

5. CONCLUSIONS

In order to reduce the unsteadiness of response and to achieve the precise and steady positioning of the rotary-linear motor with two-dimensional ultrasonic actuators, Fuzzy PI control has been tried to apply and the parameters have been modified. The obtained results are as follows.

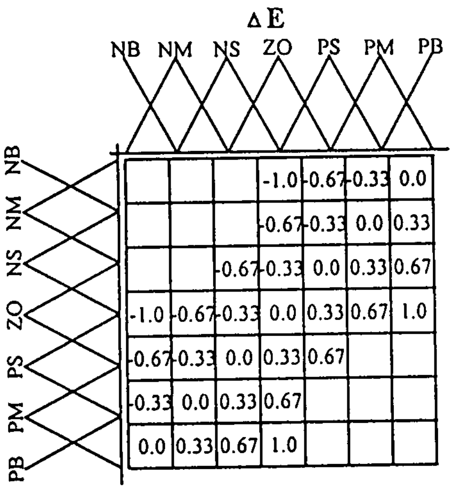
(1) Equally spaced membership functions were used and the gain parameters were adjusted based on the experimental data obtained by appling the conventional PI control. As the result, the positioning accuracy, 25.3μm, and the standard deviation, 4.4μm, were obtained; also, the unsteadiness of response was reduced.

(2) The range of ZO of membership fucntions were made narrower and some Fuzzy rules were added. As the result, the positioning error and the standard deviation were greatly improved to 3.4μm and 1.9μm, respectively.

(3) For frequency response, Fuzzy PI control was better than PI control in the frequency range of 0.5Hz to 1.Hz for both of gain and phase delay.

Table 3 Averaged values of positioning accuracy before and after modifying Fuzzy control(mm)

	Errors	Standard deviation	R.M.S. values
Fuzzy rules I	0.0253	0.0044	0.121
Fuzzy rules II	0.0034	0.0019	0.074



E \ ΔE	NB	NM	NS	ZO	PS	PM	PB
NB				-1.0	-0.67	-0.33	0.0
NM				-0.67	-0.33	0.0	0.33
NS			-0.67	-0.33	0.0	0.33	0.67
ZO	-1.0	-0.67	-0.33	0.0	0.33	0.67	1.0
PS	-0.67	-0.33	0.0	0.33	0.67		
PM	-0.33	0.0	0.33	0.67			
PB	0.0	0.33	0.67	1.0			

Fig. 11 The Fuzzy inference used for experiments of frequency response

Table 4 The adjusted input gain for frequency response

Gain	Value
G_1	1/(1.5mm)
G_2	1/(40mm/s)
G_3	60

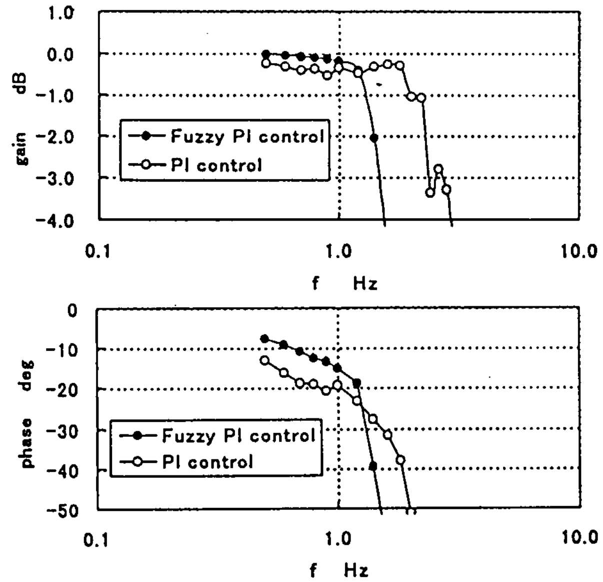


Fig. 12 Frequency responses of rotary-linear motor for linear drive

REFERENCES

[1] Nobutaka Endoh, Iwao Hayashi, Nobuyuki Iwatsuki, and Kohji Suzuki, A rotary linear motor driven by a pair of two-dimensional ultrasonic actuators, Proceedings of the 7th Symposium of Electromagnetics And Dynamics, (in Japanese), (1995), pp.431-434.

[2] Nobutaka Endoh, Nobuyuki Iwatsuki, Iwao Hayashi, Ryo Yamamoto, and Jyunji Shibata, A rotary linear motor driven by two-dimensional ultrasonic actuators, Proceedings of the 4th Japan-International SAMPE Symposium, (1995), pp.7-10.

[3] Ryo Yamamoto, Iwao Hayashi, Nobuyuki Iwatsuki, and Jyunji Shibata, Proceedings of the 8th Symposium of Electromagnetics And Dynamics, (in Japanese), (1996), pp.7-10.

[4] Ryo Yamamoto, Nobutaka Endoh, Yasuhiro Ojiro, Nobuyuki Iwatsuki, Iwao Hayashi, Jyunji Shibata, and Kohji Suzuki. The transactions of the institute of electronics, information and communication engineers A, (in Japanese), (1997), pp1736-1743.

[5] Nobuyuki Iwatsuki, Iwao Hayashi, and Yasuhiro Ojiro, Development of a two-dimensional ultrasonic actuator, Proceedings of the 2nd Japan France Congress on Mechatronics (1994), pp.763-766.

[6] Kazuo Tanaka, Advanced Fuzzy Control, (in Japanese), Kyoritsu-shuppann (1995).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronic Systems with Uncertain Physical Parameters

Erik Coelingh, Theo J.A. de Vries, Jan Holterman and Job van Amerongen

Drebbel Institute for Systems Engineering, EL-RT, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands, Phone: +31-53 489 27 07, Fax: +31-53 489 22 23
E-mail: mechatronics@rt.el.utwente.nl, www-address: http://www.rt.el.utwente.nl/mechatronics

Keywords: mechatronics, robust control, physical models, QFT, uncertain dynamic systems

During the design of controllers for mechatronic systems the designer often has to deal with uncertain physical parameters. Quantitative Feedback Theory can handle this parametric uncertainty, but this design method for robust controllers requires the uncertainty to be specified in uncertainty regions in the Nichols chart. Several techniques to map uncertain physical parameters onto uncertainty regions in the Nichols chart will be discussed. For one technique computer-based support has been developed to enhance the automated design of QFT-controllers for mechatronic systems with uncertain physical parameters

1 INTRODUCTION

Controller design for mechatronic systems generally is based on a physical model of the electromechanical plant. In the conceptual design stage as well as during operation, the *physical parameters* in the model are not known exactly or may be time varying. Lower and upper bounds for these parameters can often be indicated, resulting in so-called structured uncertainty. In the physical model of Figure 1 the uncertain physical parameters are the stiffness c and the mass of the end-effector m_1. This system is used as an example throughout this paper.

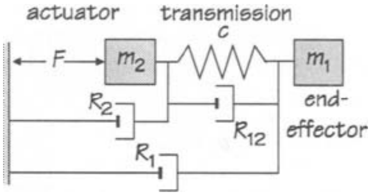


Figure 1: Typical mechatronic plant.

Often uncertainties are represented such that they lend themselves well for mathematical analysis. For robust controller design it is more appropriate to use a *physically motivated* uncertainty representation (Ackermann, 1993). To emphasize the importance of this insight, Ackermann postulates the 'first basic rule of robust control': *Require robustness of a control system only for physically motivated parameter values and not with respect to arbitrarily assumed uncertainties of the mathematical model.*

This paper addresses the problem of dealing with physical parameter uncertainty during controller design with a mechatronic design approach. First the application of Quantitative Feedback Theory (Horowitz, 1982) to this problem is motivated. Secondly several techniques are discussed that transform the description of structured uncertainty into a form suitable for Quantitative Feedback Theory (QFT). Next the techniques are evaluated for application to mechatronic systems. For one technique computer-support is developed to enhance the automated design of QFT-controllers.

2 QUANTITATIVE FEEDBACK THEORY

QFT is a design method for robust controllers that is able to deal with parametric uncertainty in a non-conservative way, while minimizing the cost of feedback (Horowitz, 1982).

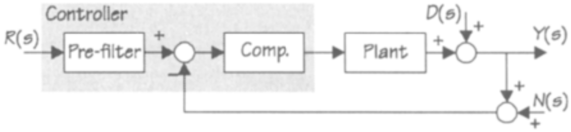


Figure 2: Feedback system (2-DOF).

Figure 2 shows the configuration of a QFT-controller that consists of a compensator and a pre-filter which will be designed consecutively.

The specifications for the controlled system have to be formulated as a desired frequency response band in a Bode diagram, e.g. as shown by the lighter shaded area in Figure 3. The bounds α and β determine the performance specification. The freedom between the bounds gives the robustness specification. For all possible uncertainties the actual frequency responses of the controlled system (B) has to lie within this specified frequency band.

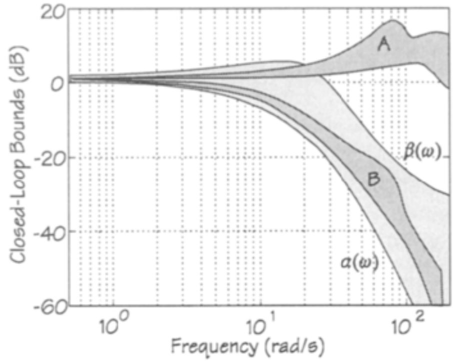


Figure 3: Frequency response bands.

The compensator is designed such that in spite of disturbances the robustness specification is met. I.e., the width of the frequency band A in Figure 3, that indicates the response of the loop for all combinations of uncertain physical parameters, has to be smaller than the width of the desired frequency response band. The pre-filter is designed such that the performance specification is met, i.e. the frequency band of the command response B is fitted into the desired response frequency band.

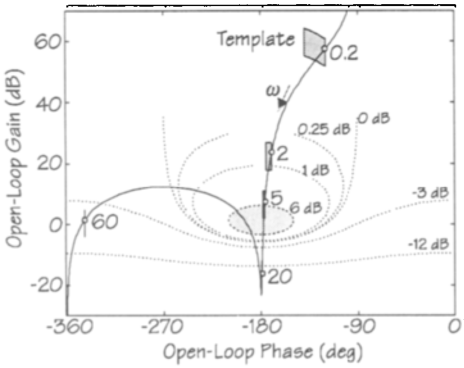


Figure 4: Templates in the Nichols chart.

Uncertainty is only dealt with for a selected set of frequencies, the so-called *critical frequencies*. For each critical frequency the structured uncertainty of the plant has to be specified as an uncertainty region in the Nichols chart, i.e. a *template* (shaded area in Figure 4). Evaluating the loop transfer function of an uncertain system for a particular frequency will result in a template in the Nichols chart, instead of one point. The contour of this template is used for loop-shaping in the Nichols chart (Horowitz, 1982).

This paper addresses the problem of constructing templates for systems with uncertain physical parameters. It is important to determine the contour of a template exactly. Over-bounding requires extra efforts to make the system robust for non-existing uncertainties. This often results in a higher bandwidth and thus a higher cost of feedback, which is undesired. The advantage of QFT to handle structured uncertainty non-conservatively is lost.

3 TEMPLATES AND VALUE SETS

The physical parameters in a plant model are gathered in a physical parameter vector $\boldsymbol{q}$. The elements of $\boldsymbol{q}$ are assumed to lie between mutually independent bounds.

$$\boldsymbol{q} = [q_1 \quad q_2 \quad \cdots \quad q_k]^T \qquad q_i \in [q_{i,-}, q_{i,+}] \tag{1}$$

In case the lower bound does not equal the upper bound, q_i is said to be *uncertain*. In the parameter space the set of possible uncertain parameter vectors is given by a hyperrectangle, the so-called *Q-box*.

$$Q = \{ \boldsymbol{q}_u = [q_1 q_2 \dots q_l]^T \mid q_i \in [q_{i,-}, q_{i,+}],\ i = 1, 2, \dots, l \} \tag{2}$$

The plant to be controlled $P(s,\boldsymbol{q})$ is a member of the parametric plant family $\mathcal{P}(s,\boldsymbol{q})$. The Q-box has to be mapped onto templates. Thus the problem of finding the template contour is replaced by finding parameter values in Q that will be mapped onto the contour. Both the numerator and the denominator of the transfer function of $\mathcal{P}(s,\boldsymbol{q})$ are polynomials in s with uncertain coefficients a_i:

$$p(s,\boldsymbol{q}) = a_0 + a_1 s + \dots + a_n s^n \tag{3}$$

The complex number taken by a polynomial when evaluated at $s = j\omega$ is called its *value*. Computing this value for each $\boldsymbol{q}_u$ in Q yields the *value set*

$$\mathcal{P}(j\omega, Q) = \{ p(j\omega, \boldsymbol{q}) \in C \mid \boldsymbol{q}_u \in Q \} \tag{4}$$

where $p(j\omega,\boldsymbol{q}_u)$ is a point in the complex plane C. A value set is a region in the complex plane that is fully determined by its contour. It can easily be transformed into a sub-template: the element (Re(p), Im(p)) in the value set corresponds to the point

$$\left(\arctan\left(\frac{\mathrm{Im}(p)}{\mathrm{Re}(p)} \right), 20\log\left(\sqrt{\mathrm{Re}^2(p) + \mathrm{Im}^2(p)} \right) \right) \tag{5}$$

in the Nichols chart. The contour of a value set is mapped onto the contour of the corresponding template in the Nichols chart. There are two classes of methods to obtain the contour of a value set (Ackermann, 1993) and hence of a template:

A priori: determine *in advance* which physical parameter combinations (possibly) yield points at the contour of the template and next calculate the corresponding phase and magnitude.

A posteriori: calculate the frequency response for all relevant parameter combinations and determine *afterwards* which phase-magnitude combinations indeed contribute to the contour of the template.

4 A PRIORI TEMPLATE DETERMINATION

An uncertain plant family can be written as the division of two uncertain polynomial families

$$P_p(j\omega_i,q,Q)=\frac{Num(j\omega_i,q,Q_N)}{Den(j\omega_i,q,Q_D)} \tag{6}$$

A template $R_p(\omega_i)$ can be determined a-priori if the numerator and the denominator of the transfer function of the plant depend on different parameters (Ackermann, 1993). In that case, the uncertain physical parameter vector $\mathbf{q}_u$ can be split in $\mathbf{q}_{u,N}\in Q_N$ for the numerator and $\mathbf{q}_{u,D}\in Q_D$ for the denominator. Figure 5 shows the procedure for a priori template determination. Each step will be discussed hereafter.

4.1 Template addition in Nichols chart

The uncertain numerator and denominator families

$$\begin{aligned} Num(j\omega_i,q,Q_N) &= \{num(j\omega_i,q_{u,N}) \mid q_{u,N}\in Q_N\} \\ Den(j\omega_i,q,Q_D) &= \{den(j\omega_i,q_{u,D}) \mid q_{u,D}\in Q_D\} \end{aligned} \tag{7}$$

can both be represented in the Nichols chart by sub-templates R_N and R_D. Template R_P can be constructed by subtracting any denominator vector from any numerator vector (Figure 6), i.e. the last step in Figure 5:

$$\begin{aligned} &\arg(P(j\omega_i,q_u) = \\ &\quad \arg(num(j\omega_i,q_{u,N}))-\arg(den(j\omega_i,q_{u,D})) \end{aligned} \tag{8}$$

$$\begin{aligned} &|P(j\omega_i,q_u)|_{dB} = \\ &\quad |num(j\omega_i,q_{u,N})|_{dB}-|den(j\omega_i,q_{u,D})|_{dB} \end{aligned} \tag{9}$$

Let n_0 be the nominal numerator in sub-template R_N and let d_0 be the nominal denominator in sub-template R_D (Figure 6*a*). If the denominator would be certain, mirrored sub-template $R_{D,mirror}$ would degenerate to nominal mirrored denominator $-d_0$. Template R_P would equal the shifted sub-template $R_N^* = R_N - d_0$, and nominal plant P_0 would correspond to the shifted nominal numerator $n_0 - d_0$ (Figure 6*b*). In case the denominator is *uncertain*, the uncertainty in the entire plant increases. The shifted sub-template R_N^* must be extended with those points that can be reached within mirrored sub-template $R_{D,mirror}$, when the nominal point $-d_0$ lies within sub-template R_N^*. R_P thus can be constructed by shifting the mirrored sub-template $R_{D,mirror}$ around sub-template R_N^*, keeping the nominal point of sub-template $R_{D,mirror}$ on the contour of R_N^* (Figure 6*c*). Template R_P (Figure 6*d*) can be regarded the result of the addition of the sub-templates R_N and $R_{D,mirror}$.

Next the problem of constructing sub-templates for uncertain polynomials, the first two steps in Figure 7, will be addressed. (Transformation of a value set into a template is performed by (5)). Depending on the type of polynomial the contour of a value set can be calculated (Ackermann, 1993). For the determination of a value set, it is important that the polynomial is classified in the simplest category.

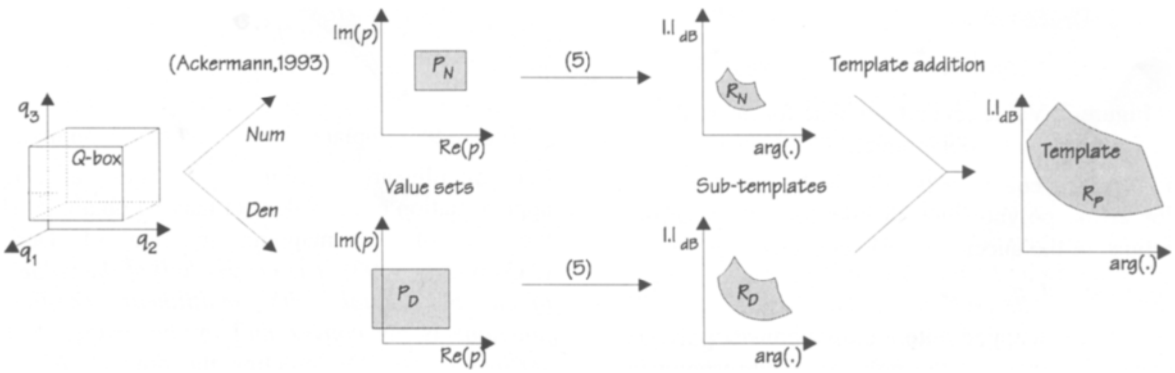


Figure 5: A priori template determination.

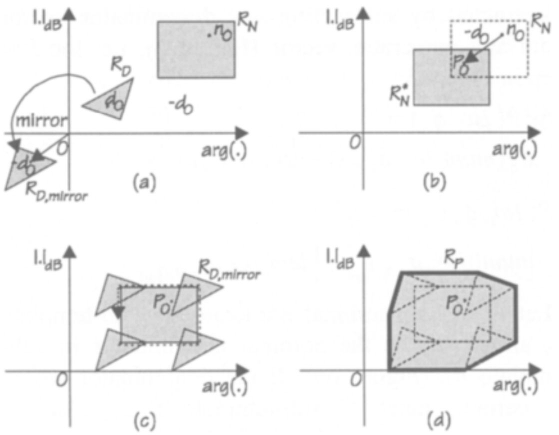


Figure 6: Template addition.

4.2 Interval coefficients

An *interval polynomial* is a polynomial for which the uncertain coefficient vector **a** ranges over the box

$$\mathsf{A} = \left\{ \boldsymbol{a} \mid a_i \in [a_{i,-}; a_{i,+}], i = 0,1,\ldots,n \right\} \tag{10}$$

For an interval polynomial a_i is not considered a function of other parameters. The contour of a value set is the rectangle between the complex points corresponding to the Kharitonov polynomials:

$$\begin{aligned} p_{+-}(s) &= a_{0,+} + a_{1,-}s + a_{2,-}s^2 + a_{3,+}s^3 + a_{4,+}s^4 \ldots \\ p_{++}(s) &= a_{0,+} + a_{1,+}s + a_{2,-}s^2 + a_{3,-}s^3 + a_{4,+}s^4 \ldots \\ p_{-+}(s) &= a_{0,-} + a_{1,+}s + a_{2,+}s^2 + a_{3,-}s^3 + a_{4,-}s^4 \ldots \\ p_{--}(s) &= a_{0,-} + a_{1,-}s + a_{2,+}s^2 + a_{3,+}s^3 + a_{4,-}s^4 \ldots \end{aligned} \tag{11}$$

This rectangle can be transformed to a template in the Nichols chart using (5), refer Figure 5.

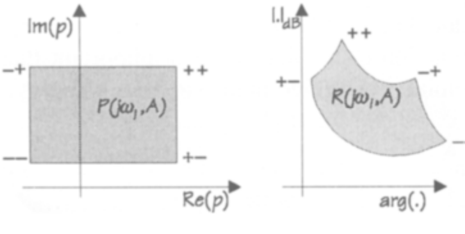


Figure 7: Value set and template for interval polynomial.

4.3 Affine coefficients

For affine polynomials all coefficients are affine functions of the uncertain parameter vector $\boldsymbol{q}_u$.

$$a_i(\boldsymbol{q}_u) = c_i + \boldsymbol{d}_i^T \boldsymbol{q}_u \tag{12}$$

The Q-box is mapped onto a closed convex polygon $R(\omega,\boldsymbol{q}_u)$. All edges of the polygon R correspond to edges of Q. The contour of the value set is obtained by calculating the polynomial value for all vertices of the parameter space Q and next determining the convex hull of these points in the complex plane.

Consider the following polynomial with uncertain parameters q_1, q_2, $q_3 \in [-1, 1]$.

$$p = q_1 + 2q_2 + (q_2 + 2q_3)s + (q_3 + 2q_1)s^2 + s^3 \tag{13}$$

The value set for ω=1 is given in Figure 8. The points marked with +++, -++ etc. correspond to vertices of Q. The edges of Q are mapped onto the straight lines between these points. The contour of the value set is the convex hull of these straight lines.

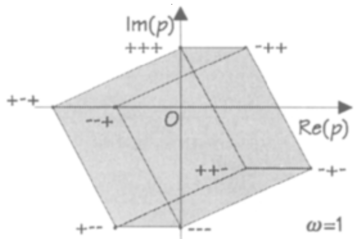


Figure 8: Template for affine polynomial.

4.4 Multilinear coefficients

Multilinear coefficient functions are functions with terms like q_1q_2, q_2q_3, $q_1q_2q_3$ etc, but no terms like q_1^2 or $q_1q_2^2$. For multilinear polynomials, the contour of the value set may correspond to extreme values of one of the physical parameters, but interior points of Q may also contribute to the contour. An illustration of this fact is given in Figure 9. The edges of Q are mapped onto the four straight lines A,B,C,D, creating the lighter shaded area. However, an interior line in the Q-box is mapped onto the curve *outside* this polygon, thus creating the darker shaded area.

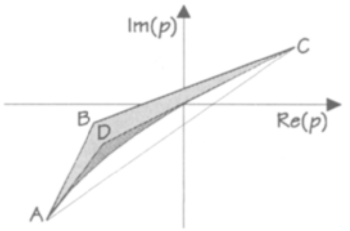


Figure 9: Template for multilinear polynomial.

For multilinear polynomial families a rough approximation of the value set may be obtained with the use of the mapping theorem of Desoer (Ackermann, 1993): *The convex hull of the value set of a polynomial with multilinear coefficient functions is the convex hull of the images of the vertices of Q.* Thus creating the polygon ABC in Figure 9, which is somewhat conservative.

4.5 Dependency numerator and denominator

For physical plants, the numerator and the denominator of a transfer function typically share dependency on some uncertain physical parameters. The a priori template contour determination method presented here will yield an over-bounded template. This is illustrated for the transfer function from input force to actuator position for the system of Figure 1:

$$
\begin{aligned}
P(s,\boldsymbol{q}) &= \frac{b_1 s + b_0}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s} \\
b_1 &= R_{12} \\
b_0 &= c \\
a_4 &= m_1 m_2 \\
a_3 &= (R_1 m_2 + R_{12} m_1 + R_{12} m_2 + R_2 m_1) \\
a_2 &= (c m_2 + c m_1 + R_1 R_2 + R_{12} R_1 + R_{12} R_2) \\
a_1 &= c(R_1 + R_2)
\end{aligned}
\tag{14}
$$

The mass of the end-effector m_1 is unknown but bounded; it may vary between 100% and 200% of its nominal value. The stiffness c may vary between 50% and 100% of its nominal value. Other physical parameters are assumed to be certain.

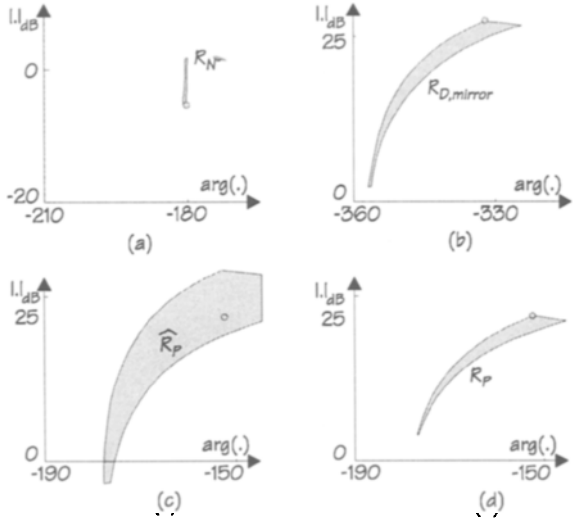


Figure 10: a) shifted numerator sub-template
b) denominator sub-template
c) overbounded template
d) actual template

The uncertain parameters occur in both the numerator and the denominator. Figure 10*a* shows the sub-template R_N of the numerator. The mirrored sub-template $R_{D,mirror}$ of the denominator, with the nominal point set to the origin, is shown in Figure 10*b*. Addition of the templates results in the over-bounded template of Figure 10*c*. The actual template R_P is much smaller, as can be seen from Figure 10*d*. In this case application of a priori template determination (Figure 7) results in over-bounding.

5 A POSTERIORI TEMPLATE DETERMINATION

The contour of a template can also be determined a posteriori. The response ($\arg(P(j\omega,\boldsymbol{q}))$, $|P(j\omega,\boldsymbol{q})|_{dB}$) can be calculated for a finite set of parameter vectors in the *Q*-box. This subset is referred to as the *grid* and the approximation technique is referred to as *gridding* (Ackermann, 1993). The problem in gridding is to determine the density of the grid beforehand. Visual inspection of the template however yields an immediate answer to the question whether an applied grid was dense enough or not. Therefore it is advisable to start with a relatively low grid density that can be iteratively increased (Figure 11 and Figure 12). Once the gridding result is dense enough one has to determine the points of the template that contribute to the contour.

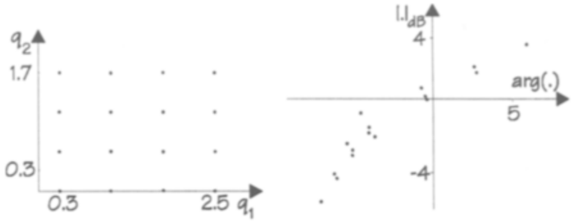


Figure 11: Gridded Q-box and gridding result.

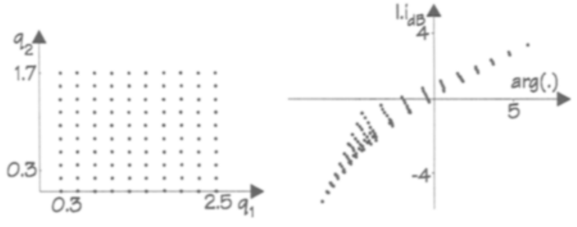


Figure 12: A denser grid, a better gridding result.

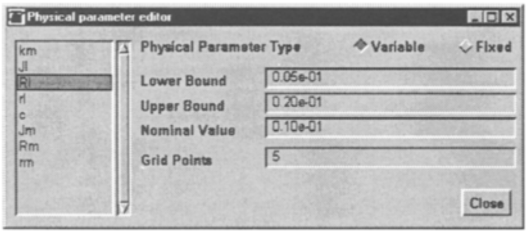
6 TEMPLATES FOR MECHATRONIC SYSTEMS

A priori determination of the contour of a template is only applicable for a small group of models. In case the uncertainty shows complexity it is difficult to determine beforehand which physical parameter combinations contribute to the template contour. For this reason many controller design problems do not take parametric uncertainty into account quantitatively. A typical mechatronic design problem would be to make the system of Figure 1 robust for load variations and stiffness uncertainty. Even for the relatively simple transfer function (14) it is not straightforward to determine the template contours a priori, as the numerator and denominator are not mutually independent. The stiffness occurs in both polynomials.

Gridding is a straightforward, maybe time-consuming, but nearly errorless method that can easily be automated. The *a posteriori* method will be used to enhance the automated design of QFT-controllers for mechatronic systems with uncertain physical parameters. An interesting feature is that the designer can select grid-points on the contour of the template and investigate which combination of parameter values corresponds to that particular point. In this way, the mechatronic designer is given the opportunity to signal physical parameter uncertainties that unnecessarily complicate controller design. The designer could reconsider the plant design in order to avoid undesired situations.

7 COMPUTER-BASED SUPPORT

Computer-based support has been developed that automatically converts a physical plant model with uncertain physical parameters into templates. First the physical plant model is built in 20-sim (Weustink et al., 1998). Physical parameter values are entered as a number or as a range, as in Figure 13.



Physical Parameter Type ◆ Variable ◇ Fixed

Lower Bound 0.05e-01

Upper Bound 0.20e-01

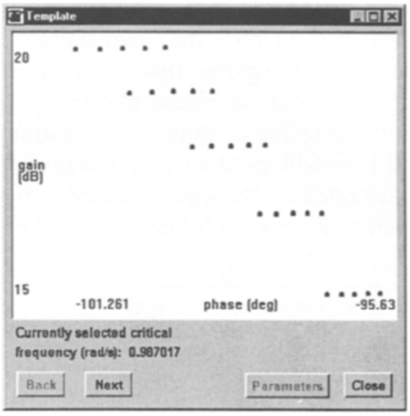
Nominal Value 0.10e-01

Grid Points 5

Figure 13:The physical parameter editor.

For uncertain parameters the number of grid points have to be specified. A symbolic plant model is exported to Matlab together with the data concerning the physical parameters. A special purpose M-file calculates the frequency responses for the specified grid, and draws the templates using the Matlab QFT toolbox (Borghesani et al., 1994). Figure 14 shows a template for a specific frequency.

The designer should indicate the contour of the template that is needed for the QFT design. The QFT-toolbox (Borghesani et al., 1994) can now be used for controller design.



Currently selected critical frequency (rad/s): 0.987017

Figure 14: The template editor.

8 CONCLUSIONS

Controllers for mechatronic systems should be robust for physical parameter uncertainty without a high cost of feedback. With Quantitative Feedback Theory these controllers can be designed, but uncertain physical parameters have to be mapped onto templates in the Nichols chart. Methods for a priori template determination depend on the type of polynomials of the plant transfer function and can only be used when uncertain physical parameters do not occur in both the numerator and denominator. A posteriori template determination (gridding) is a straightforward and nearly errorless method that can be applied to all transfer functions. This technique is the most appropriate for mechatronic systems. Computer-based support has been developed that automatically grids the parameter space of a linear time-invariant plant model and that automatically determines the templates. These can be used by the Matlab QFT-toolbox. The designer is given the opportunity to signal physical parameter uncertainty that unnecessarily complicates controller design.

9 REFERENCES

Ackermann, J., (1993), *Robust Control, Systems with Uncertain Physical Parameters*, Springer Verlag, London, U.K.

Borghesani, C., Y. Chait,, O. Yaniv, (1994), *Quantitative Feedback Theory*, The Mathworks. Inc., Natick, MA, U.S.A..

Horowitz, I.M., (1982), Quantitative Feedback Theory, *Proceedings of IEE*, Part D, **129**(6), pp. 215-226.

Weustink, P.B.T., T.J.A. De Vries and P.C. Breedveld (1998), Object-Oriented Modeling and Simulation for Mechatronic Systems with 20-sim, submitted to Mechatronics '98.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

The Design and Implementation of a Surface Orientated Controller

A.J. Winsby, S.E. Burge, I.A. Grout

Engineering Department, Lancaster University, Lancaster, LA1 4YR, United Kingdom

This paper introduces a novel approach to implementing control strategies by consideration of the input-output controller surface. The approach is intended to allow the realisation of control schemes based on classical methods together with more recent approaches such as fuzzy logic. The design tools and procedures necessary to implement a surface controller are outlined with reference to a simple case study.

1. Introduction

Recent research at Lancaster [1] has proposed a novel approach to implementing control strategies by consideration of the input-output controller surface. This approach allows the realisation of schemes based on classical methods together with more recent approaches such as fuzzy logic [2]. To facilitate implementation of this novel control method an Application Specific Integrated Circuit (ASIC) has been designed and fabricated, a software version of this ASIC has been written and an integrated controller design/implementation platform developed. This paper considers the design tools and flow necessary to produce a control system based on the surface approach. At all stages of development, cost and minimisation of complexity to the user have been major considerations. This paper begins with a statement of the controller problem that outlines the surface approach. The design flow and tools are then introduced and illustrated with reference to a simple case study. The whole process crosses many traditional subject boundaries which fits well into the Mechatronic ethos. This is reflected in the conclusion section which also compares the results of the case study and outlines the direction of present and future research into the surface approach to controller design and synthesis.

2. Statement of Controller Problem

Many control strategies are currently available to the control engineer for the solution of control problems. The procedure in applying any particular strategy is broadly similar: analysis of plant, determination of controller structure, implementation and tuning. For the case of conventional PID (proportional + integral + derivative) type schemes, the implementation may be via analogue electronics, software or ASIC [3]. For Fuzzy Logic the implementation is most likely to be software based although some hardware devices have been both proposed and developed [4]. Each strategy generally requires a range of differing skills, devices, development platforms and methods. However by taking the surface approach a number of different strategies can be implemented using a single device.

Recent research at Lancaster [1] has proposed the concepts of Surface Orientated Control (SOC) and Direct Surface Generation (DSG). DSG attempts to provide a common point of reference for the large number of control methods/strategies currently available to the control engineer whilst SOC seeks to provide a controller structure which is able to support the implementation of many different control strategies using the same basic processing algorithm. DSG and SOC have arisen from a multidisciplinary approach to the solution of engineering problems, which allows a common coherent framework to be developed for integrating conventional control strategies with more recent advances such as fuzzy logic.

2.1 DSG and SOC Defined

The driving principles behind DSG and SOC can be summed up in the following three statements:

1) If an Ideal (or close to ideal) control action exists for any particular control problem, then regardless of

which control method is employed (i.e. fuzzy logic, conventional, neural networks etc.), then given the same set of inputs the resulting controller will possess the same input/output surface.
2) If a method exists for approximating to a good degree of accuracy almost any function (linear or non-linear), then any controller input/output relationship and hence any control action can be produced.
3) A device which is capable of supporting multiple control methods is potentially more useful than one which is restricted to a single method.

In essence, any control strategy may be used to produce a controller surface. This permits a novel approach to controller analysis and synthesis. Moreover it also permits implementation via either software or hardware. To facilitate hardware implementation an Application Specific Integrated Circuit (ASIC) has been designed and fabricated and a software version of this ASIC written, both of which form part of an integrated development/implementation system.

2.2 The Surface: storing and reconstructing a controller input-output surface.

The controller surface (or surfaces) provide the transformation between the controller input(s) and output(s). This transformation may be linear or non-linear. In the current scheme surface data is stored as a 33x33 matrix of data points. These are restored to a continuous controller surface by interpolation between adjacent matrix values as required within the surface controller.

This paper considers controller surfaces with two inputs and a single output. It should be noted however that the method is by no means limited to controllers of this type. The rest of this paper will introduce the Development Platform and the design flow involved in creating, simulating and implementing a surface controller. The process is illustrated by considering both a conventional theory defined surface and a fuzzy logic defined surface as applied to a DC motor position controller. The resulting surface based control systems are then compared and contrasted and conclusions drawn.

3. Development Platform Requirements

If the validity of the surface representation is accepted, then it becomes necessary to identify the major requirements of a suitable surface development platform. These can be summarised as follows:

1) Minimisation of complexity to the user.
2) Surface design capability.
3) Simulation capability.
4) Compatibility with current control methods.
5) Coherent design flow.
6) Computational efficiency.
7) Adaptability and flexibility.
8) Efficient storage and retrieval of surface data.
9) Minimisation of costs.

Although it would be possible to write suitable software to fulfil the above requirements, such a task would be complex, expensive and tantamount to 'reinventing the wheel'. The preferred route is to use existing software which has the advantage of user familiarity and quality assurance. Matlab [5] is particularly suitable with its collection of control toolboxes such as Fuzzy logic and Neural Networks, together with its simulation and visualisation capabilities. The matrix representation fundamental to Matlab allows fast and convenient storage, manipulation and analysis of surface data. In essence, Matlab is used as the design/simulation environment whereby all the normal techniques can be used to design for example a conventional controller or a fuzzy controller. The end product is however not a set of gains or a rule base, but a surface (or surfaces) which can then be used with an ASIC model in Simulink [6] to simulate the control system.

A software version of the surface controller ASIC has been written in C. This can be called from the Matlab environment and used to test the designed controller via a commercially available PC based A/D card. The card is also used for data acquisition enabling the controlled system to be monitored and the resulting data to be analysed within the Matlab environment.

When a suitable control system has been designed and tested, a choice of implementation methods is available:

1) The resulting surface data is loaded into a data storage EPROM and the surface controller ASIC is able to take control of the system.
2) Using the software version of the ASIC, a PC based control system can be implemented.
3) The software can be directed toward a dedicated Microprocessor/Microcontroller based system.

4. Design Flow and Case Study

The design flow can be broken down into three distinct processes. These are illustrated in Figure 1. As with most control system design, the process is iterative, often requiring many passes through each process to achieve an acceptable controller design. These three processes will be further illustrated with reference to a simple case study. This being the closed loop position control of a DC electric motor module.

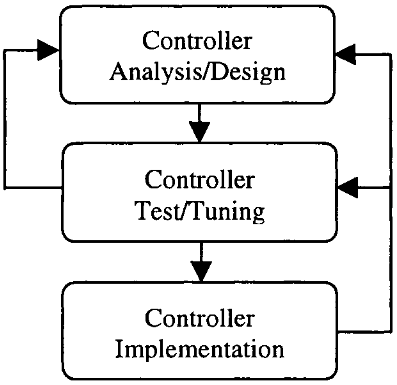


Figure 1: Surface Controller Design Flow

4.1. DC Motor Module Description

The DC motor module consists of a 12V DC motor housed in a module containing interfacing electronics, position sensing potentiometer and tachogenerator. The outputs from the module are therefore motor position (0^0 to 360^0, -5V to +5V) and motor speed (-5V to +5V, 0V= stationary). The single input to the module is the motor drive voltage input (-5V to +5V).

The target ASIC and software controller also have associated signal conditioning circuitry to level shift and attenuate inputs to a 0V to +5V range.

Figure 2 shows the motor module and controller in closed loop position control configuration.

5. Controller Analysis/Design

Two methods of controller design are considered in this study.

1) Using the classical approach based upon the mathematical analysis of the DC motor module yields the algebraic PV (proportional + velocity) control law:

$$c(t) = K_p e(t) - K_v v(t)$$

where $c(t)$ = controller output
$e(t) = x(t) - y(t)$ = positional error
$v(t)$ = velocity feedback
K_p = proportional gain
K_v = velocity gain

The surface equivalent of this control law is shown in Figure 3.

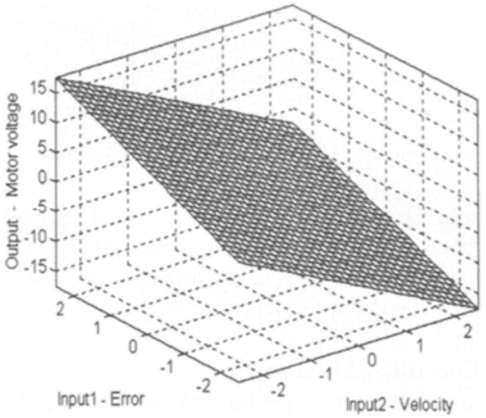


Figure 3: Conventional Controller Surface

The surface slope relative to the *error* and *velocity* inputs are equal to the values of K_p and K_v in the PV control law.

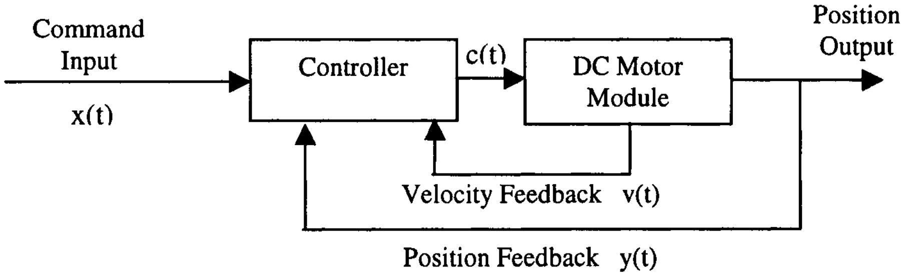


Figure 2: Closed Loop Motor Position Control

2) Using a linguistic description of the desired operation of the motor module, a fuzzy logic controller can be designed. The Matlab Fuzzy Logic toolbox allows for several different methods of inference and defuzzification. The surface shown in Figure 4 has been extracted from a fuzzy controller developed using Mamdani Inference, centre of gravity defuzzification and a rule base consisting of four rules.

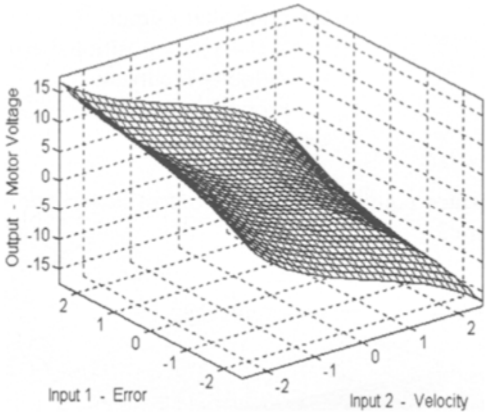


Figure 4: Fuzzy logic derived Surface

It should be noted here that no special techniques have been applied to generate these surfaces. They arise naturally as a consequence of the two controller design approaches.

5. 1. Controller Simulation

All simulations are performed within the block diagram simulation environment of Simulink. A model of the ASIC/software controller has been developed and can be included in a simulation as a controller block. Simulations can be performed to determine:

1) If the controller designs in their original form (fuzzy or conventional) achieve the desired system performance.
2) If the surface controllers produce the same results as the original equivalent controllers.

To test 1) and 2) above, three different modes of simulation may be performed.

1) Simulation of the original controller design with a plant model.
2) Simulation of the surface controller with a plant model.
3) If no plant model is available then random inputs can be generated and applied to the surface controller and original controller in parallel. Correlation of their outputs should give an indication of equivalence.

If the performance of the surface controller is found to be satisfactory then it can be tested against the real plant. If the performance is not satisfactory then two choices are available:

1) Return to the controller design phase of development, make the necessary changes an re-simulate.
2) Alternatively, operate on the surfaces directly, this will be illustrated in the next section.

6. Controller Test/Tuning

The second distinct process in the surface controller design flow is that of controller test/tuning. This is achieved within the surface controller development system by the provision of a software version of the ASIC controller written in the C programming language. This program communicates with the outside world via a commercially available PC based analogue/digital IO (Input/Output) card. The card offers 16 channels of analogue input, 4 channels of analogue output, three independent timers and the ability to generate interrupts. Combined with the controller software is a simple data acquisition scheme by which unused IO card analogue inputs can be used to monitor signals within the control system.

Surface data prepared in Matlab is read into the program when initialised in the form of a text file. Data acquired during the operation of the software controller is sent to the Matlab working directory upon exit of the control program in the form of a binary data file. This data can then be analysed and plotted within the Matlab environment to determine the controllers performance.

Figure 5 shows such a plot, this being of the step input response of the DC motor module using the conventionally defined surface shown in Figure 3. The response is slightly over-damped with zero steady state error. This response is identical to that obtained when controlling the DC motor module with a conventional PV control program using the gains $K_p=3$, $K_v=4$.

Figure 6 shows a plot of the step input response of the DC motor module using the surface controller with the fuzzy logic defined surface.

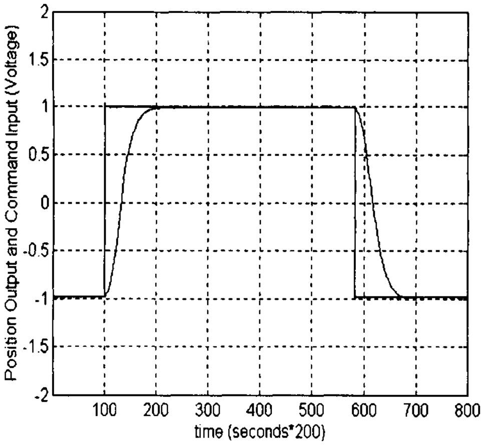


Figure 5: Conventional Surface Step Input Response

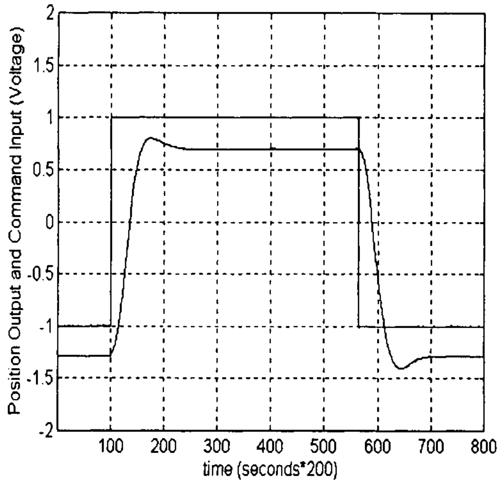


Figure 6: Initial Fuzzy Surface Step Response.

This response is inadequate in that there is a steady state error of 0.32 volts in the measured signals corresponding to the step input and the motor position output. There is also an overshoot in both the positive and negative step directions.

Normally this would necessitate a return to the first procedure in the controller design flow. Parameters would be adjusted within the fuzzy controller design and simulation/test performed until satisfactory results were obtained.
An alternative to this is to operate directly on the surface itself. This can be done by considering the step response and relating it to the surface characteristics.

1) Eliminating the steady state error:
normally with this type of controller it is desirable that when error=0 and velocity=0 then the controller output should fall to zero. Checking the fuzzy surface revealed that the controller output = 0 when the velocity = 0 and the error = 0.32volts. This accounts for the steady state error of 0.32volts. To remove this error the whole surface must be shifted relative to the output axis so that the output = 0 when the error = 0 and the velocity = 0.

2) Eliminating the overshoot:
the existence of overshoot suggests that the control system damping needs to be increased. In terms of a conventional PV controller this would be achieved by increasing the velocity gain K_V. In terms of the surface equivalent of the fuzzy controller the same effect can be achieved by increasing the slope of the surface relative to the velocity axis.

Figure 7 shows the step input response of the fuzzy surface controller after tuning by the direct surface manipulation detailed above. The response shows critical damping and zero steady state error.

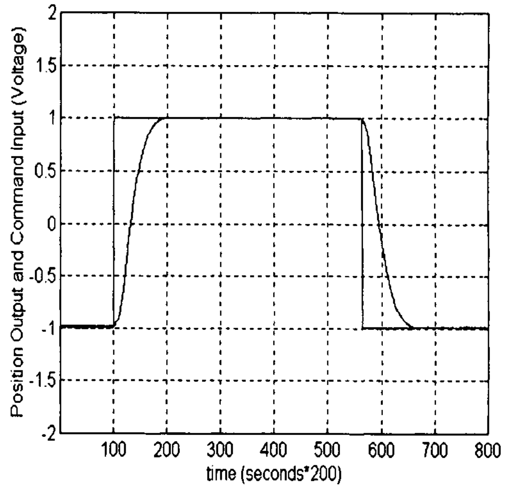


Figure 7: Tuned Fuzzy Surface Step response

6.1 Surface manipulation Toolbox

As part of the development system a collection of Matlab functions has been created with which to manipulate surfaces. These functions typically involve: shifting, rotating, scaling, and the smoothing of surfaces. The functions may act on the surface as an entity or on individual data elements or blocks of elements within that surface. The use of some of these tools in controller tuning was demonstrated above. It is recognised that direct surface manipulation offers some interesting possibilities in controller design and synthesis. These will be reflected upon in the conclusion section.

7. Controller Implementation.

Once surface development has been completed and satisfactory controller action has been established, the controller can be implemented by any of three methods, two of which have been realised and one of which is subject to future development.

1) The surface data is programmed into an EPROM and the surface controller ASIC with suitable interfacing electronics is able to take control of the system.
2) The software version of the ASIC can be used within a PC based control system, either as a stand alone program or integrated into a larger control scheme. Surface data is read from either a text file or a binary data file when the system is initialised.
3) Optimising the surface controller software should enable it to be targeted towards microprocessor or microcontroller based applications. This method of implementation is still subject to development.

One of the perceived advantages of the surface approach is that the processing algorithm used within both ASIC and software is largely independent of the method used to define the controller surface. This allows a wide range of surface types to be implemented with a minimum of reconfiguration of either software or hardware.

8. Conclusions

This paper has introduced a novel method of designing and implementing controllers. Moreover, by taking a multidisciplinary approach (including microelectronics, control, programming, data acquisition and analysis), it has been possible to produce a complete development environment using where possible existing tools and techniques.

Cost reduction, minimisation of complexity to the user, adaptability and flexibility have been maintained by adopting Matlab as the core element of the development system. The facility to create user defined functions within Matlab has enabled a *Surface Manipulation Toolbox* to be developed which allows a surface generated by any particular control method to be tuned or manipulated without having to return to the original method of design. Current research is considering hybrid surface controllers in which several different control methods are combined to produce controller surfaces.

The results of the case study show that similar performance has been obtained from both the conventional and fuzzy logic defined surfaces. The marginal differences in performance can be explained in terms of the relative slopes of the two surfaces. The surface representation also allows the two completely different control approaches to be compared and analysed using a common point of reference.

References

[1] Winsby, A.J. Direct Surface Generation and Surface Orientated Control. Introduction, Motivation, Application, Implementation. Lancaster University Internal Report, 1996.

[2] Y, J. Langari, R. Zadeh. Industrial Applications of Fuzzy Logic and Intelligent Systems. IEEE Press. ISBN 0-7803-1048-9, 1995

[3] Grout, I. Burge, S. Dorey, A.. 1994 Application Specific Integrated Circuit Implementation of SISO Control Laws. 4th International Conference on Control, Proceedings of, pp1104-1110, March 1994

[4] American Neuralogix Inc, Fuzzy MicroController Development System.

[5] MATLAB, The Math Works Inc. USA

[6] SIMULINK, The Math Works Inc. USA

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronics for Moving Machinery - a precision, multi-axis, high speed and acceleration winch system

N. H. Darracott[a] and M. Honeywill[b]

[a]DCA Design Consultants, 19 Church Street, Warwick, CV34 4AB, UK*

[b]Unusual Rigging Limited, The Wharf, Bugbrooke, NN7 3QD, UK

Integrating disciplines and the application of mechatronic techniques can achieve novel solutions which overcome previous constraints. A mechanical power transmission application for a lifting system in the field of entertainment engineering is described, incorporating AC motors and drives, high efficiency interchangeable gears, fail-safe operation and rope tension management. An auxiliary system electronically geared to the main lifting motor automatically eliminates rope slack and migration on the rope drum, using embedded processing within the drive. The successful development of a soft-electro-mechanism to improve the dynamic performance of wire rope winches is discussed and some other areas with potential for mechatronic solutions are proposed.

1. INTRODUCTION

The extreme demands of the entertainment industry create some unusual opportunities for engineering innovation, for example where the appeal and impact of live performance has to compete with the realism achieved by artificially generated effects. In both staged and filmed events there is a need for controlled motion of equipment and personnel which can be intuitively directed, programmed and re-programmed quickly, rehearsed and repeated over and over to co-ordinate with other activities and then performed to create the desired spontaneous emotion in the audience. To "catch the moment" requires the integration of many disciplines, increasingly dependent on the successful application of techniques more usually associated with advanced engineering in academia or industry.

A key to the successful application of automated motion systems in live performance applications is the effective combination of appropriate software and hardware to give adequate dynamic characteristics and meet the user-interface requirements peculiar to this sector of the entertainment industry. Sufficient "soft" dynamic performance can normally be achieved by PID control, and can be provided by the application of standard industrial products. The need for simultaneous multi axis control is not readily met by industrial control systems providing sequential process control. This coupled with the degree of interaction with an automation system that is required within this sector of the entertainment business has necessitated the development of bespoke user-interfaces and micro-processor based control systems. The evolution of variable speed drives for standard electric motors, in particular the capability to control speed and torque output from standstill, has facilitated the adoption of electrically powered machinery for lifting applications. This controllability can only be fully exploited when combined with accurate feedback and efficient mechanical elements, for example absolute optical encoders for positional feedback, precision gears and high performance belt drives for efficient and quiet operation. Other constraints traditionally found within the entertainment industry, though now increasingly prevalent elsewhere, are the extreme

---

[*] previously at Unusual Rigging Limited

short lead times and inherent need for flexibility in product functionality - it is intrinsic within this creative business that decisions are made late and change often. Standardisation and modularisation - of both components and design approach, and the employment of flexible manufacturing techniques help to deliver within these lead times.

2. BACKGROUND

Computer control for mechanical handling is long established and has developed along with the technology, for example mechatronics has been applied in various specialist applications [1&2]. Similarly mechanical aids to moving scenery in live performance have been used for many years [3], the applications advancing with available equipment, though the technological progress has not been as significant as that in computer animation and effects generation. The scope for dramatic effect and visual stimulation is to some extent limited by the technical capabilities of the equipment and systems available; as well as the safety imperatives, particularly in relation to moving people. Software based control systems are now widely applied in the field of entertainment engineering, with bespoke operator interfaces tailored to the needs of real time programming and rehearsal [4]. Elsewhere the development of electronic drives for AC and DC motor control has been applied to many applications in mechanical handling industry. More recently the development of software-configurable drives for electric motors and other system tools has broadened the performance horizons, highlighting previously non-critical limitations. For example, the speed and acceleration of wire rope winches has until now been restricted by the dynamics of conventional machines. A particular problem arises when wishing to rapidly accelerate loads suspended on a flexible cable, to create the effect of free fall whilst maintaining precise control and repeatability. Developments in control software have enabled easy programming, repetition and, theoretically, high dynamic performance. However, these high accelerations of the same order of magnitude as g are translated into the rotational acceleration of a rope drum and the tension in the rope due to the resultant force is insufficient to overcome the inertia of the rope. This causes loops to form in the rope as it feeds off the drum and allows slack rope to slip round the drum, migrating in the opposite direction to rope travel. The result is a dangerous loss of control, high shock loads (in addition to the high forces required to achieve the desired acceleration) and damage to rope and machine. The development of a solution is the subject of this paper.

3. ENGINEERING

The main mechanical elements of the primary (winching) and auxiliary (tensioning) axes are shown in figure 1. For the primary axis, a standard squirrel cage induction AC brake-motor was chosen as the prime mover. This offers the advantage of electronic drive, integral fail-safe brake, low cost, ready availability, simplicity and cleanliness of power source. A special is that the motor must provide the full effort required to support the load when stationary prior to acceleration and after deceleration. The transition at zero speed must be smooth and finely controllable. This is achievable with closed loop vector control, and by first instructing the drive to hold the motor at zero speed, providing whatever torque is demanded, then releasing the brake; the reverse is required when stopping.

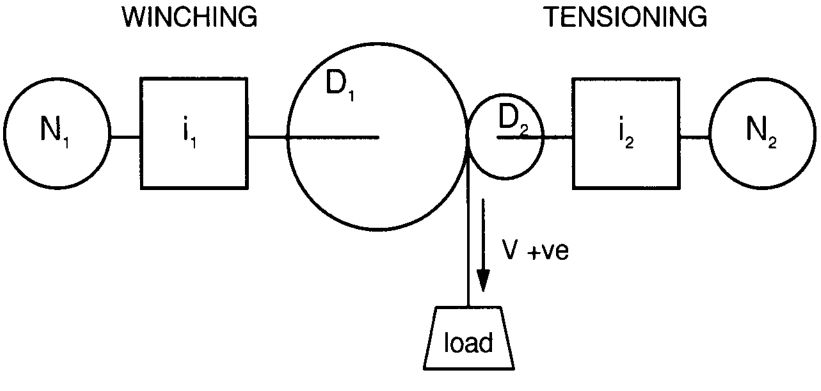


FIGURE 1

Early testing of this highlighted the importance of automatically ensuring that the brake was only released if the drive was operational, and protecting the drive from tripping when the brake was energised. As mentioned above positional feedback is important, here measured at the rope drum using a rotational encoder. The relationship between the rotation of the drum and the length of suspended rope is critical to the accuracy, for this reason the rope is wound in a single layer in a machined groove at a constant diameter.

The gearbox selected was a 2-stage helical gear unit. The modular machined casing provides an envelope which can accommodate different ratios, the gearbox can therefore be interchanged to create a different speed / torque band for the winch - for example changing from a rating of 1000 kg at 1 m/s to 200 kg at 5 m/s. Further advantage was made of the gearbox, for mounting ancillary components such as the position feedback system and as an integral part of the chassis. In order to ensure safety for personnel, a secondary fail-safe brake is required. To reduce the size and power requirements of this device a second gearbox is incorporated as a torque converter, allowing use of a brake similar to that integral with the primary motor. An additional benefit is derived by using this gearbox as the second bearing and support for the winch drum, eliminating much of the precision fabrication, machining, and assembly of a traditional machine chassis. The costs and associated lead time are thus eliminated, replaced by sourcing a standardised product. The structure of the winch was developed from this, such that the primary alignment and longitudinal stiffness is derived from the gearboxes and drum. These are mounted on a minimal sub-frame along with auxiliary components. The assembly is bolted into a lightweight tubular space-frame which provides torsional stiffness, enclosure and flexibility for installation.

One way of overcoming the problem of slack when rapidly paying rope off the drum is to provide a roller in contact with the rope, at the tangent point of the rope and the drum, and to drive this roller such that it tends to tighten the wrap. This roller is driven by an auxiliary motor, also a standard squirrel cage induction AC motor, via a toothed belt drive. The roller is mounted in a swinging frame, pivoting about the auxiliary motor axis. The frame is prevented from oscillating by a large spring force, independent of the small radial force between roller and rope drum which can be readily adjusted. The roller is constructed from thin wall steel tube, with stud shafts welded to either end and post machined, this construction minimises inertia and maximise stiffness. The circumference of the roller is surfaced with a adhesively bonded friction material.

For zero slip between rope, drum and roller $V = V_1 = V_2$ where V, V_1 and V_2 are speeds of rope, drum circumference and roller circumference respectively. It is required that $V=V_1$, however if $V_2 - V_1 > 0$ there will always be a positive (tightening) force F acting on the rope, determined by $F = \mu_d . R$ where μ_d is the dynamic coefficient of friction between rope and roller surface and R is the clamping force between roller and drum. By linking the two axes electronically, the relationship between them can be altered according to the direction of travel, thus enabling $V_2 - V_1 > 0$ (i.e. always positive) for V_1 positive and negative. This relationship is subject to the functionality offered by the selected drive, in this case it is possible for $V_2 = v + s . V_1$ where v is offset, s ratio. Expressed in terms of rotational speed this becomes:

$N_2 = n + s . k . N_1$

where $k = D_1 . i_2 / D_2 . i_1$ (the ratio between N_1 and N_2 required for $V_1 = V_2$)

and, for $N_1 \geq 0$, $s = a \geq 1$, $n = c > 0$

for $N_1 < 0$, $s = b < 1$, $n = 0$

k, a, b and c are variables in the auxiliary drive.

The main electronic and control elements are summarised in the simplified system diagram, figure 2. The simple case illustrated comprises two axes of motion (two identical winches of the type described above), common applications can contain 200 axes, typically operating with simultaneous control of more than 20 axes. The PC-based computer control system provides all the functionality required by the user for configuring, programming, and operating the motion system and is linked by a data network to an interface for each axis. This interface includes the motion controller for the axis, with absolute position feedback from the mechanics. The encoder used has integral signal conditioning and is synchronously driven from the rope drum. The primary motor drive operates in closed loop vector mode with the drive physically remote from the motor. The auxiliary motor drive also operates is closed loop vector mode, but autonomously from the user control system and axis motion controller. A speed signal from the primary motor feedback loop is used as a reference for the auxiliary motor and the motion control is achieved by imbedded software within the auxiliary motor drive. Values for the variables k, a, b and c were determined by experimentation. Utility software run on a laptop computer allowed easy adjustment during testing to optimise performance according to speed/torque band of the winch, anticipated acceleration, roller pressure, materials, rope diameter and construction.

A third axis within the machine is mechanically driven from one end of the winch drum. This is a spooling mechanism to provide rope fleet management and incorporates a belt drive, leadscrew, linear slideway, structural beam and modular rope sheave assembly. Further mechanical and electrical elements not described here, are incorporated in the winch to provide other functional and safety features. Thus "one axis" contains two axes of motor control, linked with a switchable characteristic "electronic gearbox", 3 encoders, 4 mechanical axes, 6 mechanical transmissions.

This apparent complexity requires some justification, and this is found in the application and performance requirements. One example is a desire for people to appear to drop in free-fall, traditionally achieved in the entertainment industry through the use of stunt artists and carefully staged scenes. A limitation of this is the ability to repeat the motion precisely and many times without prolonged re-setting. A need arose to achieve this for two people together, to be able to repeat the motion in different locations and to interface with computer effects and graphics which required defined controllability of acceleration and speed. The PC-based supervisory computer control system, provides not just the user interface, but also the ability to group axes together - synchronously, in harmony or sequentially; plotting, modifying, rehearsing, interrupting, reversing and running cues; and to stop / start safely in emergency or fault conditions.

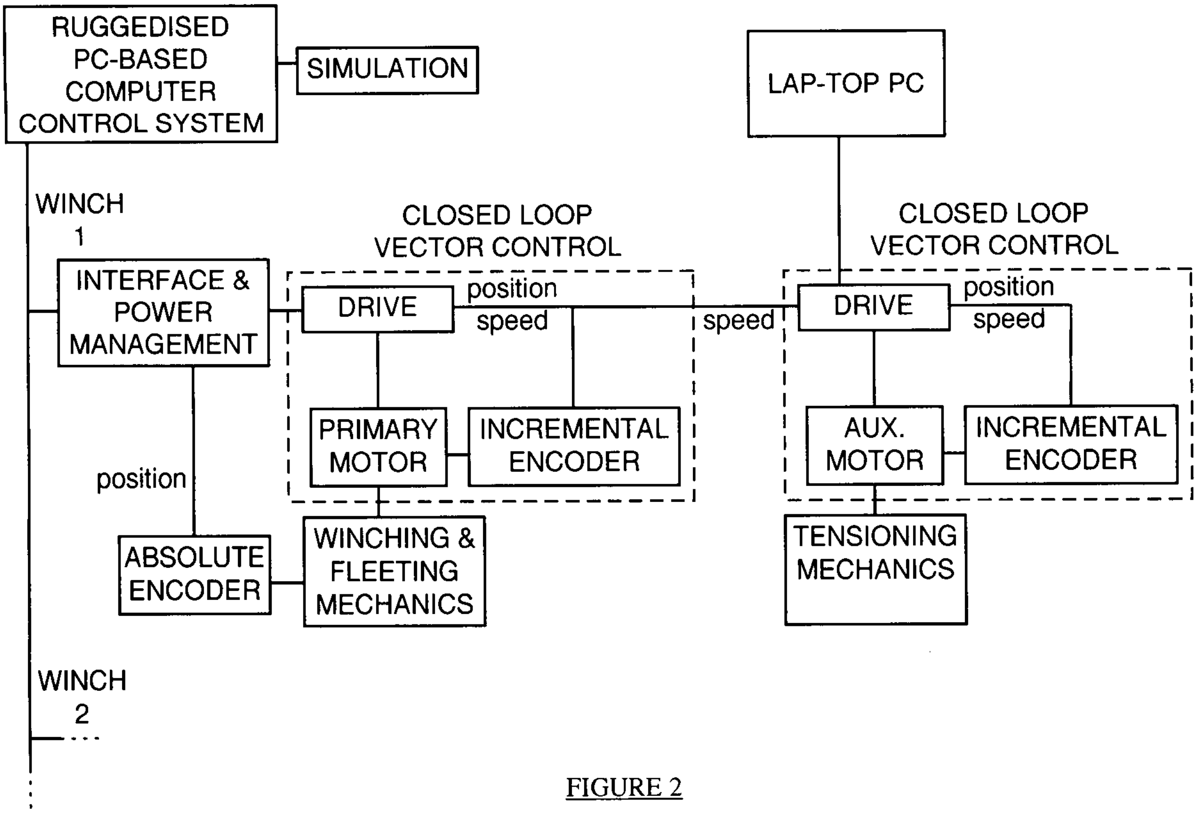

RUGGEDISED PC-BASED COMPUTER CONTROL SYSTEM
SIMULATION
LAP-TOP PC
WINCH 1
CLOSED LOOP VECTOR CONTROL
CLOSED LOOP VECTOR CONTROL
INTERFACE & POWER MANAGEMENT
DRIVE
position
speed
speed
DRIVE
position
speed
PRIMARY MOTOR
INCREMENTAL ENCODER
AUX. MOTOR
INCREMENTAL ENCODER
position
ABSOLUTE ENCODER
WINCHING & FLEETING MECHANICS
TENSIONING MECHANICS
WINCH 2

FIGURE 2

4. DESIGN

As so often the successful development of this novel system was dependent on more than theory and the application of technology [5]. A modular design strategy was developed, segregating areas of the system, not by traditional disciplinary boundaries, but by logical function and perceived end value. This allowed parallel progress, the use of previously developed components / sub-systems, limited the complexity and risk of each design stage and was key to the fast-track design and build program. A requirement of this type of approach is that the problem must be sufficiently constrained early on, in order to permit division. This needs a sound basis of technical understanding and practical knowledge in order to have confidence in "drawing the line" in the right place, or at least satisfactorily. A rigorous methodology was applied, which can be considered as comprising two stages of activity. These were: (i) identifying a) principles, b) an ideal solution, c) known technologies, d) applicable and available new technologies; (ii) development leading to a) analysis of the proposed solution and its performance and modularity implications, b) checking against first principles and experience, c) common sense - the appropriate application of knowledge possessed, d) recording calculations and decisions. Two aspects consistently maintained were (I) a meticulous attention to detail, aided by CAD assembly of parts during design, robust design, and an organic parts list - structured and developing with the progress of design and (ii) broad vision of the project objective - incorporating issues such as the effect, safety, time and cost. The simplification of components and assemblies, through manufacturing techniques and standard parts facilitated the very fast development time and the maintainability and adaptability of the resulting system. The high performance demands, whilst presenting the major technical problem also provided the goal which drove development, without such demands there would have been no progress. Whilst many recognisable design principles [6] were employed, the development work ventured beyond the traditional boundaries of mechatronics to create a comprehensively multi-disciplined activity, incorporating structures, mechanics, electrics, electronics, software, human factors and logistics - the whole system had to be transported by airfreight and fit through an office doorway.

5. CASE STUDY

The nature of the project to provide a solution for the filmed stunt application described above was such that the guarantee of a successful outcome within a very tight time constraint was commercially essential. It was therefore decided to initially pursue two potential solutions to the problem of eliminating slack rope accumulation and migration on the rope drum, one mechanical, one mechatronic. Both were investigated until the balance of demonstrable advantages against disadvantages for one solution sufficiently outweighed that of the other to provide confidence in attaining the required goal. In hindsight it is possible to treat this as a case study of the real benefits that can be offered by mechatronics over other disciplines. Some of the key factors are summarised in table 1, though not exhaustive this list gives a useful insight into the decision made and associated reasons for adopting a mechatronic solution to this highly constrained drive transmission problem.

Table 1

	Nature of solution	
advantage (dis-advantage)	Mechanical	Mechatronic
theoretical	simple. (fixed ratio). (direction).	variable ratio. (multi-disciplinary).
practical	(limited flexibility). (sourcing). (existing constraints).	adaptable end result. experience. (complex).
commercial	cost.	knowledge development. future use. (risk). (reliability).

Though in theory a mechanical solution would have been relatively simple, the requirement to have a different relationship between axes according to the direction of rotation would have added to the actual complexity. The consequences of an inherently fixed ratio would have required this to be determined early on in the design stage and without the benefit of experimentation, the ability to readily vary the relationship in the mechatronic solution overcame this and provided an additional commercial benefit of increased flexibility for future use. Multi-disciplinary understanding is needed when adopting mechatronics and can prohibit such a solution; however the experience, necessary skills and working relationships were in place to overcome this potential difficulty. Surprisingly the practical disadvantages of a mechanical solution were unmitigated; major concerns were the inherently limited flexibility in design and end result; the difficulty of sourcing components in the limited time available and need to fall within existing constraints of the application. In contrast, the mechatronic solution promised an adaptable end result with more "headroom" to develop the solution within the known constraints. Though intrinsically complex, previous successful experience with mechatronic drive systems had demonstrated the practical advantages and straightforward operation of systems, once designed and commissioned. One of few advantages of the mechanical solution was its perceived lower cost, though realistic costings were significantly higher than initial budget estimates and much closer to the mechatronic system than expected. Combined with longer term commercial benefits of the mechatronic solution, which included increasing company knowledge and experience as well as greater potential for future adaptation and other uses, the mechanical solution became less attractive. The potential risks of the mechatronic system failing to perform were acknowledged and judged manageable, as mentioned above this was partly achieved by initially advancing on two fronts, until sufficient confidence was established. Further concern was raised over the system reliability, which was addressed by a very simple failure mode and effects analysis and steps taken to enable rapid fault finding and easy serviceability.

6. CONCLUSION

The specialised application for which the work described here was undertaken focused on a need to ensure improved dynamic performance of a wire rope winch, previously limited by the tension in the rope being insufficient unwind from the drum during high speed / rapid acceleration. The development of a mechatronic solution achieved the objectives and has opened up possibilities for further refinements, including combined variable torque and speed characteristics. There are other limitations of current winching systems that might also lend themselves to a mechatronic solution, two examples are the need to only coil rope in a single layer in order to provide an accurate position feedback information and the requirement for a mechanical device to keep the rope fleet angle within acceptable limits. More general applications include the dynamics of stopping in emergency situations whilst providing duality and optimising the physical bulk and performance of moving machinery. These and others provide some interesting opportunities for engineering development and demonstrate that the surreal world of movie making challenges the mainstream of engineering and other disciplines to provide real working solutions.

REFERENCES

1) Clist G., Harwood D. &Williams J., "The Use of a Gyroscopic Device to Control Suspended Loads", Proc. 5th UK Mechatronics Forum Int. Conf., Vol. 1, pp 157-160, Guimares, Portugal, 1996

2) Lee D.W., Hesterman D.C., Mace B.R. & Jones R.W., "Dynamic Modelling of Hydraulic Machinery for Estimation Purposes", Proc. 5th UK Mechatronics Forum Int. Conf., Vol. 1, pp 415-420, Guimares, Portugal, 1996

3) Finney M, "Pick That Thing Up!, Motorized and Automated Rigging Systems for Leisure Attractions", Proc. TiLE '94 International Conference on Technology in Leisure and Entertainment, pp 70-80, Maastricht, The Netherlands, June 1994

4) Darracott N.H. & Douglas J., "Moving Moving Lights - The Automated Motion Control of Lighting Equipment and Other Visual Effects", Showlight '93 International Colloquium Film, Theatre and Television Lighting, Bradford, England, April 1993

5) Roberts R.E.J., "Management, Engineering and Innovation", Proc. Instn Mech Engrs Vol 203, pp 1-15, 1994

6) French M.J., "An Annotated List of Design Principles", Proc. Instn Mech Engrs Vol 208, pp 229-234, 1994

ACKNOWLEDGEMENTS

All work was performed in the course of normal business, whilst employed by Unusual Rigging Limited. The support and encouragement of the company and the effort of many individuals is gratefully acknowledged, their contributions merit recognition.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronics in medical engineering: Advanced control of a ventilation device

Imad Jenayeh, Frank Simon and Heinrich Rake

Institute of Automatic Control,
Aachen University of Technology, 52056 Aachen, Germany
Phone: ++49-241-807481, Fax: ++49-241-8888296,
E-mail: je@irt.rwth-aachen.de

A typical mechatronic system consists of a mechanical process, electromechanical actuators, electronic sensors and a controller unit with the corresponding software. In this paper a microcontroller-based digital feedback control of a positioning device for a ventilation machine is presented. This kind of machine should allow either a volume- or a pressure-based controlled ventilation. The interdependence between the two important physiological variables, respiration volume and lung pressure, is highlighted and the requirements on the control system are discussed. Based on a mathematical model for the respiration device and the characteristics of the patient lung a feedback control concept for both respiration modes has been developed and is presented in this paper. By means of simulations the control concept has been tested. To sum up some simulation and experimental results are shown.

1. INTRODUCTION

In a co-operation between the Institute of Automatic Control and the Dräger AG, Lübeck, a well-known developer and manufacturer of ventilation devices, a new generation of respiration devices has been developed. This new, cost-effective machine generation will give the operators more comfort and better functionality and the developers more flexibility. The presented solution for flexible and safe artificial respiration is an extension of a control concept which was developed and implemented in a former co-operation.

A comparison between natural and artificial respiration shows that artificial respiration needs sophisticated control systems. Natural inhalation is reached by elastical expansion of the rib cage. Because of the depression an air flow fills the lung with fresh air. The following contraction of the rib cage causes exhalation. Therefore, inhalation is an active process where muscle work is needed whereas exhalation is passive. The imitation of this natural process by a technical device is possible, but causes a high technical expenditure. The realisation of artificial respiration becomes much easier if the pressure course is inverted. Inhalation causes an overpressure because fresh air is actively pumped into the lung and exhalation is realised by evacuation. This inversion of the natural circumstances requires careful operation of the ventilation device and supervision of the critical values lung pressure and respiration volume.

2. STRUCTURE OF THE VENTILATION DEVICE

The artificial ventilation is principally based on periodically filling and evacuating the lung by means of a controlled moving up and down of a piston.

Fig. 1 shows the basic assembly of the ventilator. It consists mainly of the piston which is moved by means of a DC motor. An incremental encoder gives back the actual position of the piston.

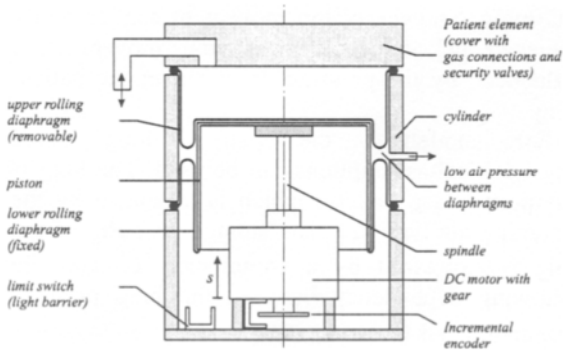

Patient element (cover with gas connections and security valves)
upper rolling diaphragm (removable)
cylinder
low air pressure between diaphragms
piston
lower rolling diaphragm (fixed)
spindle
s
limit switch (light barrier)
DC motor with gear
Incremental encoder

Fig. 1 Scheme of the ventilator

The basic structure of the control system of this ventilation device is shown in Fig. 2 . The available signals for control purposes are the armature current, the piston position resp. the piston speed and the piston pressure. A so-called master controller is responsible for the user interface, monitoring tasks and for sending the desired parameters via a serial interface. The second controller executes the control algorithm in real time. The actuating variable (armature voltage) is generated by the power electronics and depends on a pulse width modulation signal (PWM) generated by the microcontroller.

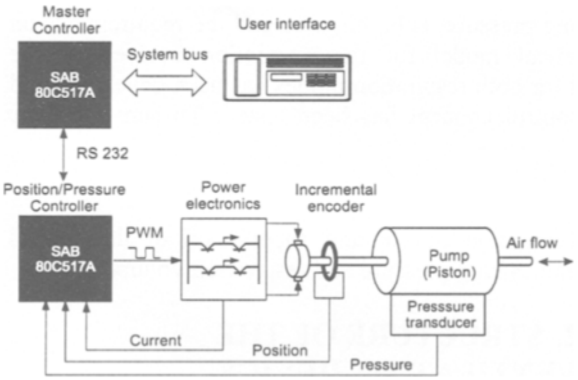


Fig. 2 Basic structure of the control application

This structure gives full flexibility and functionality to the system. Some mechanical solutions could be transferred into software which emphasises the mechatronics aspect of this device.

3. MODELLING

The description of the dynamic behaviour of the DC motor and the piston movement can be easily achieved [1], whereby the piston movement is influenced by the pressure produced in the patient lung.

For modelling the patient lung some physiological assumptions can be used. The lung is interpreted as a balloon which is expanded by the incoming air flow [2]. The storage property of the lung is expressed by a compliance C and the following dependency between the lung pressure p_C and the air flow $\dot{v}_p$

$$p_C = \frac{1}{C}\int \dot{v}_P dt \,. \tag{1}$$

The air pressure acting on the piston includes furthermore a second part p_R caused by the flow and the resistance R of the respiratory tracts

$$p_R = R \cdot \dot{v}_P \,. \tag{2}$$

The superposition of both effects leads to a typical course of the lung pressure and the corresponding air volume (see Fig. 3), whereby the interdependence between both variables has to be considered.

During the inspiration flow phase a determined air volume is pumped into the patient lung and correspondingly the air pressure increases rapidly until the first pause is reached. During this pause some settlement processes occur in the lung. During the expiration phase a similar procedure takes place in the opposite direction.

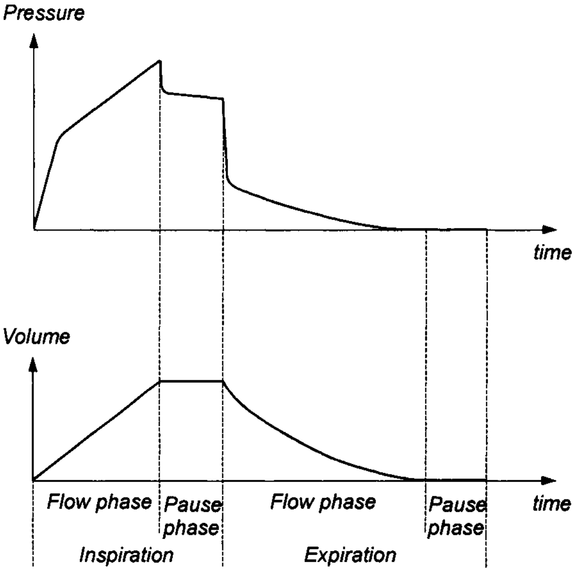


Fig. 3 Typical ventilation cycle

Fig. 3 also shows a difficulty of artificial respiration. The lung pressure and the inspired volume are coupled by the specific properties of the patient lung. From a control point of view it would be desirable to manipulate the flow and the pressure independently. Unfortunately, this is quite impossible because there is only one manipulated variable, the armature voltage. Thus a controller could be designed either to follow a position profile, determining air volume and flow or a pressure profile. By prescribing the air volume and flow the resulting pressure course is unpredictable because of varying patient parameters R and C. Especially the pressure peak at the end of the inspiration phase can be harmful for the patient lung. On the other hand prescribing a pressure profile is also problematical because the necessary flow and volume to reach the pressure are variable. In this case a minimum volume of fresh air in every respiration cycle is essential. Of course, the interdependence cannot be cancelled by any kind of control but the critical bounds can be supervised and in the case of violation the operation mode must be modified automatically to ensure the patient safety.

For controller design and simulation purposes the obtained model is implemented by means of the graphic-, block-oriented simulation package MATLAB®/SIMULINK® which presents a good platform to test the different control strategies which will be discussed in the next section.

4. CONTROL CONCEPT AND DESIGN

As mentioned above there are two important variables to control, lung pressure and respiration volume. Therefore, two basic operation modes are possible:

- volume-based ventilation with supervised pressure (IPPV[1])
- pressure-based ventilation with supervised volume (PCV[2])

Both operation modes should be realised including the already existing control of piston position and speed.

During the volume-based ventilation the bounds for the lung pressure are set by the operator and must be respected to guarantee a safe ventilation. This has been solved up to now by means of some additional safety valves in the respiratory device to influence the lung pressure and finally to keep it within the desired bounds. The disadvantage of this solution is that the desired volume for ventilation is lost. This leads to the extension of the existing air volume feedback control with a pressure control loop.

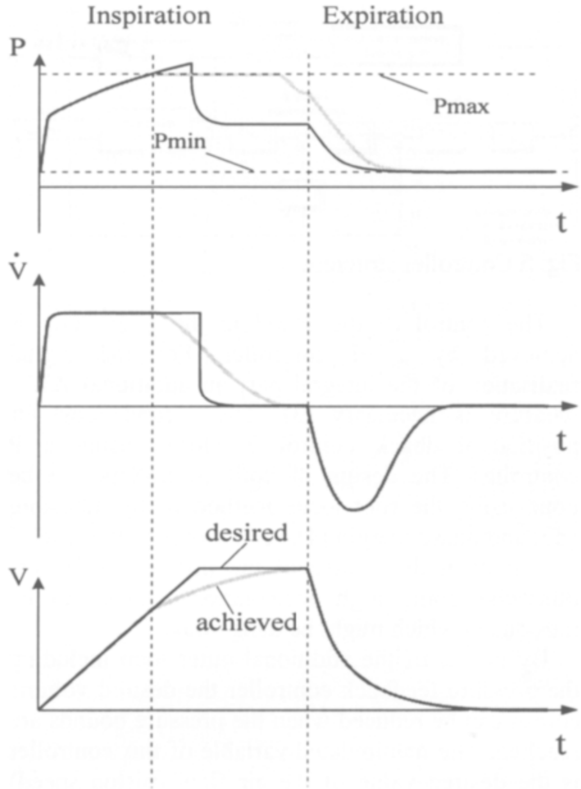


Fig. 4 Desired operation of the control system

Fig. 4 shows an example of the desired behaviour of feedback controller (volume-based ventilation) with and without a bound for the pressure in the lung. A desired reference course of the air volume can be tracked until the upper pre-set bound is reached. The pressure controller is switched on to maintain this pressure level. Consequently, the air flow is reduced and the upper piston position (final respiration volume) is reached (if possible) later. In analogy with this the same task is repeated during the expiration phase (not shown in Fig. 4) where the lower bound has to be respected and maintained.

As a result of the extension of the previous, already existing volume control with an additional pressure control loop, a pressure-controlled ventilation can be achieved whereby a pressure profile is used instead of a fixed setpoint (P_{MIN}/P_{MAX}). In this case, it is necessary to bound the air flow.

The controller concept is based on the structure shown in Fig. 5 . The inner loop of the presented cascade structure controls the air volume (piston position) and the air flow (piston speed).

---

[1] IPPV Intermittent Positive Pressure Ventilation [2]

[2] PCV Pressure Controlled Ventilation

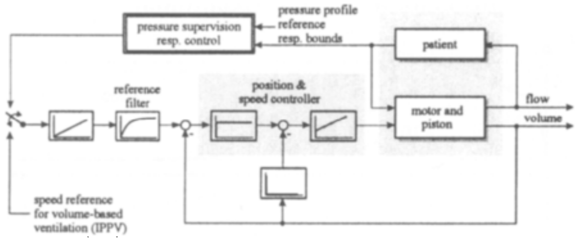

pressure supervision resp. control
pressure profile reference resp. bounds
patient
reference filter
position & speed controller
motor and piston
flow
volume
speed reference for volume-based ventilation (IPPV)

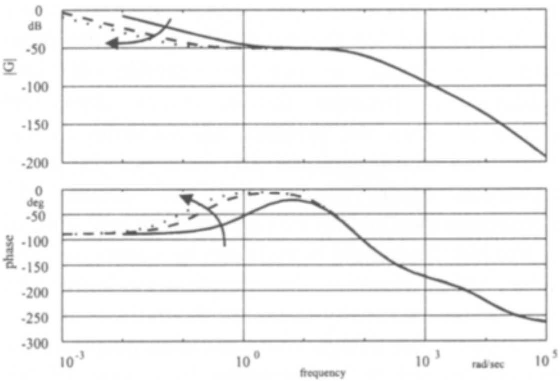
Fig. 5 Controller structure

The control of the speed in the inner loop is achieved by a PI controller. For the digital realisation of the integral part an additional ARW measure is necessary [3]. The second loop for position feedback control is closed using a P controller. The design of both controllers can be done using the root locus method or by adjusting gain and phase margin [4]. A first-order filter is used to smoothen the course of the position reference, otherwise some high acceleration values can be introduced which might be dangerous.

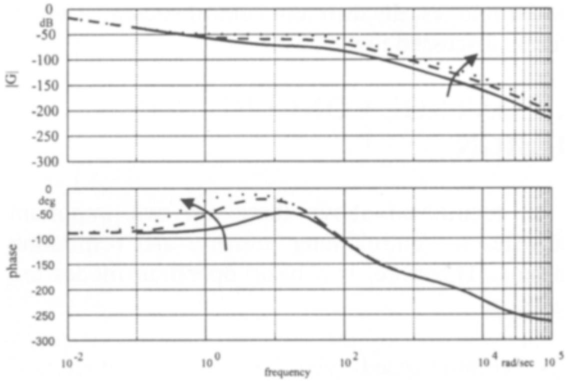
By means of the additional outer loop including the pressure feedback controller the desired volume course can be reduced when the pressure bounds are reached. The manipulated variable of this controller is the desired value of the air flow (piston speed) which has an integral effect with time delay on the air pressure. Therefore, a PD controller is the best choice to compensate possible pressure deviations.

For design of the pressure controller the strong influence of the patient parameters *R* and *C* on the dynamic behaviour of the controlled system must be considered. As the ventilation device should operate for a variety of patients from infants to adults the parameters change in a wide range. Fig. 6 shows the open loop frequency responses for typical values of *R* and *C*.

Increasing values of *C* (see Fig. 6a) lead to an increasing phase and a decreasing magnitude for lower frequencies. However controller design isn't influenced by the compliance as there are no differences of magnitude in the frequency range of interest. Things are different for a variation of the resistance. There is a considerable difference in the magnitude for higher frequencies which can cause stability problems. To ensure a reliable and robust operation of the ventilation device an adaptation of the controller parameters subject to the different patient characteristics is indispensable. In a first step this problem has been solved by defining 4 classes for typical patients and an appropriate controller was developed. Further efforts could be spent in the on-line identification of the patients and the self-tuning of the control parameters.



a) frequency response for different values of *C*



b) frequency response for different values of *R*

Fig. 6 Influence of *R* and *C* on the open loop

5. RESULTS

Fig. 7 shows an example for a pressure-supervised, volume-based ventilation. The simulation result outlines the intended behaviour of the pressure feedback control that is switched on when the upper pressure level is reached. The pressure controller slows down the volume flow in order to keep the pressure level until the inspiration phase attains the end or the upper desired piston position is reached. The same concept works for the lower pressure level which is sometimes needed in respiration physiology. Analogously, the pressure-controlled ventilation tries to follow a given pressure profile but simultaneously the resulting air flow must be bounded.

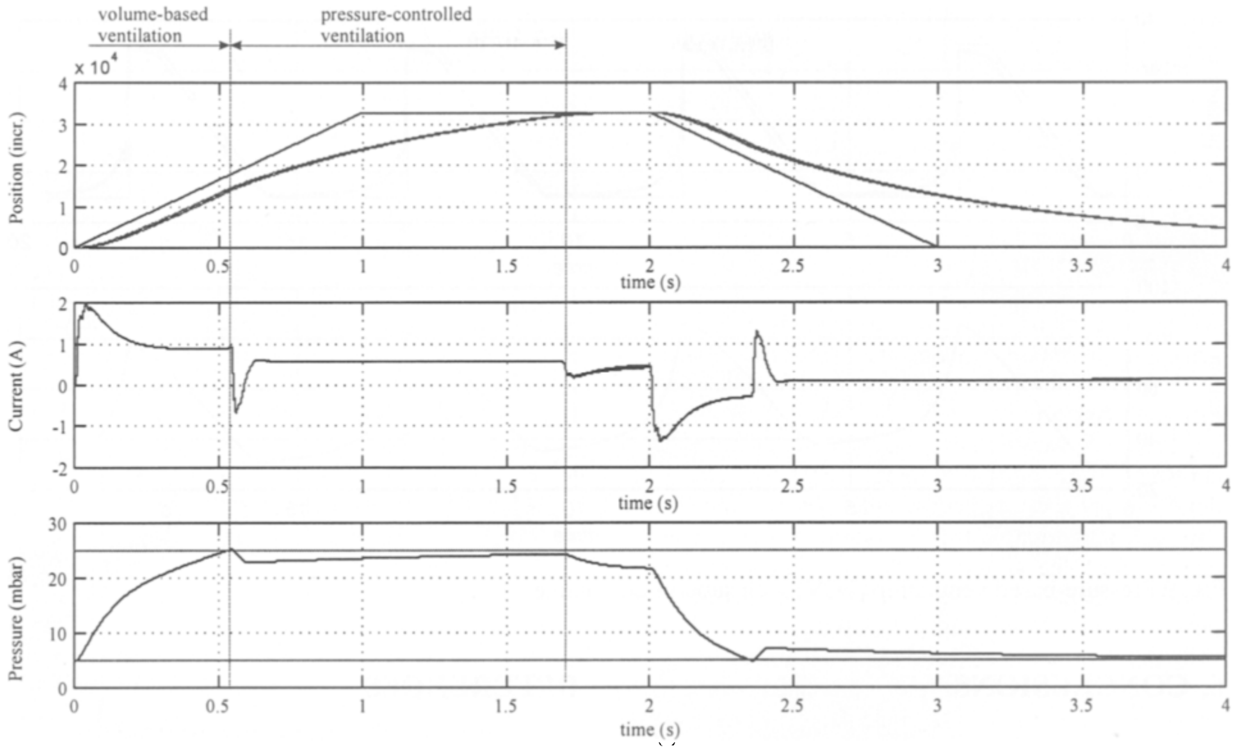


Fig. 7 Simualtion results for IPPV mode

The implemented control strategy has delivered the experimental results shown in Fig. 8 and Fig. 9.

Fig. 8 emphasises the improvement reached after introducing the pressure feedback control in the IPPV mode. In the first two phases one can see the typical course of the pressure while the given (trapezoidal) volume profile is tracked. In phase III the lower bound was reached and the pressure feedback control loop is switched on. The piston speed changes to keep the pressure within the desired bounds. Analogously, in IV the upper pressure bound is kept till the end of the inspiration phase. Fig. 9 shows the new ventilation mode which is based on the tracking of pre-set pressure profiles whereby the air flow must be kept within patient-dependent bounds.

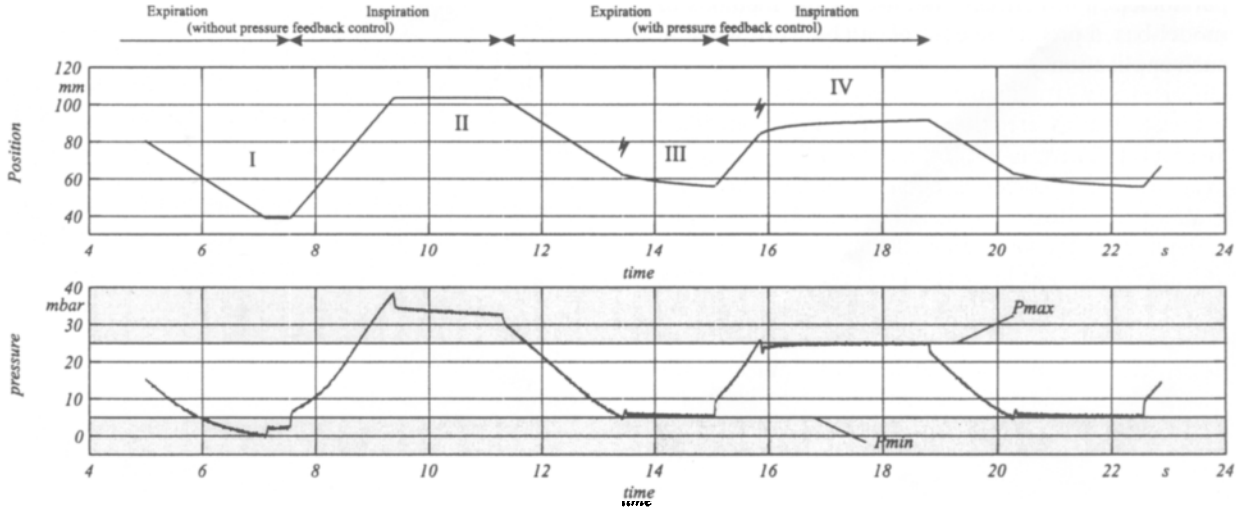


Fig. 8 Volume-based ventilation (IPPV) with supervised pressure

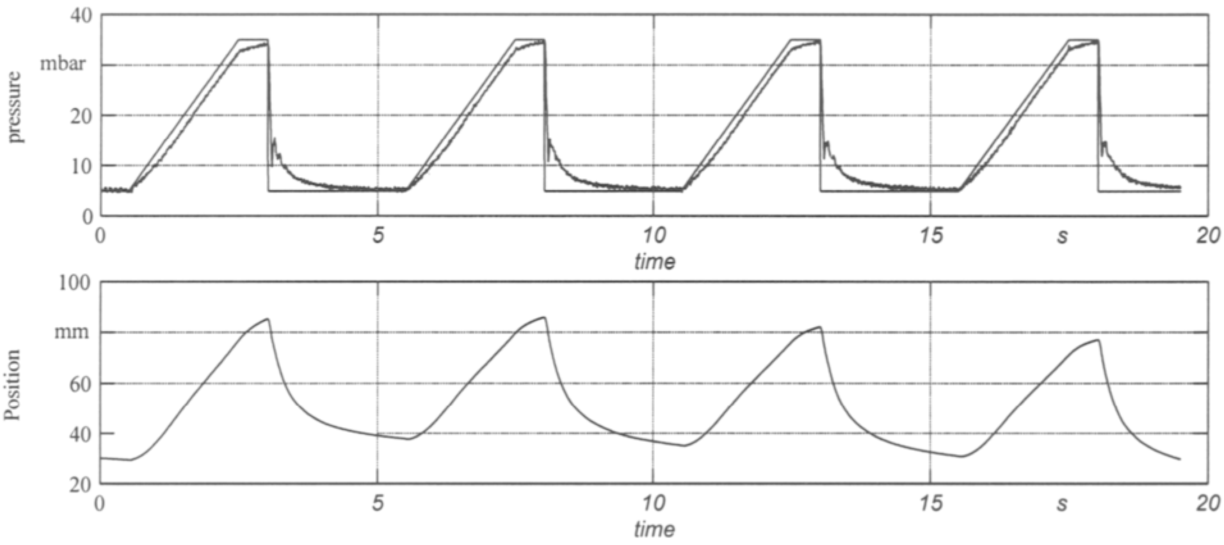


Fig. 9 Pressure-based ventilation (PCV) with supervised volume

6. CONCLUSIONS

The presented control concept increases the safety in the artificial ventilation by means of control and supervision of the very important physiological variables air volume and air pressure. The behaviour of the ventilation device has been improved and a new ventilation mode (PCV) has been introduced. The automatic pressure supervision and control will relieve the operator. By means of simulation results the control concept has been checked and was successfully implemented in a prototype of the ventilation device.

Future efforts will be spent to improve the robustness of the controller by on-line identification of the patients and automatic tuning of the controller parameters. Furthermore, the use of the methods of model-based predictive control can be worthwhile in such applications.

REFERENCES

1. I. Jenayeh, F. Simon, S. Bernhard, H. Rake and B. Schaible, Digital Control of a Positioning Device for a Ventilation Machine. European Control Conference ‘97, Volume 4 ,(1997)
2. Lotz, P., Siegel, E., Spilker, D., Grundbegriffe der Beatmung, GIT Verlag Darmstadt, 1994.
3. Noisser, R., Anti Reset Windup measures for single loop controls with digital controllers, at, 12, pp. 299 - 304. (1987)
4. Rake, H., Regelungstechnik A, Script RWTH-Aachen, 20 Edition, (1997)

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Modeling of A Distributed, Non-Steady State, Thermal System For The Purpose of Control

Michael M. Chen and Dr. Kevin C. Craig
Department of Mechanical Engineering, Aeronautical Engineering and Mechanics
Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA

Dr. Gerald A. Domoto
Mechanical Engineering Sciences Laboratory, Xerox Corporation
141 Webber Avenue, North Tarrytown, NY 10591, USA

This paper discusses the issues involved in using Finite Element Method (FEM) to model a transparent xerographic fuser roll as a distributed, non-steady state, thermal system for the purpose of control design. Matlab® and Simulink® are used as primary simulation tools due to their wide applications in the field of control system designs.

1 Introduction

During the xerographic process, fusing is a very important stage where the toned image is bonded to the substrate permanently. The most common fusing method is to apply heat and pressure so that the toner particles meld and adhere to the paper. The conventional xerographic fuser uses rubber coated hollow metal rolls which are heated with lamps from inside. Due to the thermal mass of the rolls and the way they are heated (from inside to outside), it usually takes a long time for the roll surface to warm up to the operating temperature.

In order to reduce the warm-up time, transparent fusers have been invented. Figure 1 shows one of the transparent fuser configurations. The roller is a hollow transparent Pyrex glass cylinder coated with black Teflon. Unlike a traditional fuser roller, in this configuration, the thermal radiation from a heating lamp inside the hollow cylinder transmits through the glass and heats up the outer boundary of the roller first. This enables the fast heat up of the fuser surface. Thus, the transparent fuser has an instant-on capability, which not only reduces the warm-up time but also eliminates the energy consumption required to maintain the stand-by temperature. Hence, the configuration is much more energy efficient. However, since the transparent roller has less thermal mass and faster dynamics, it requires a more sophisticated control scheme to regulate the fusing temperature.

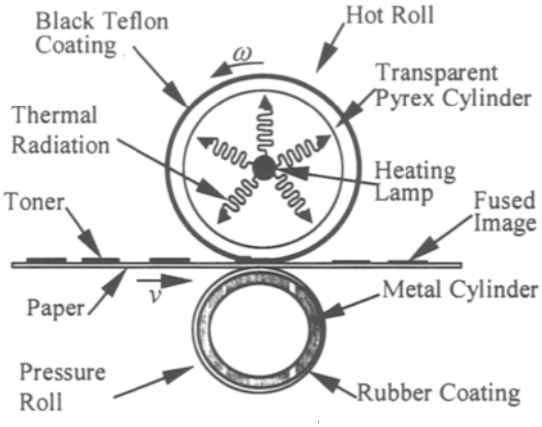


Figure 1. Transparent xerographic fuser schematic

One very important step in the mechatronic approach to control system design is to obtain a model that sufficiently represents the plant to be controlled. This paper discusses the issues involved in using FEM to model the transparent fuser roll as a distributed, non-steady state, thermal system for future control design purposes.

2 Problem Statement

The general transient temperature distribution inside the Pyrex cylinder is an initial boundary value problem which is governed by the following partial differential equation

$$\rho C_p \frac{\partial u}{\partial t} - k\Delta u = Q \tag{1}$$

where u is the temperature distribution within the glass roll, Δ is the Laplace operator, k is the thermal conductivity, ρ is the density, C_p is the specific heat, and Q is the heat source.

To simplify the problem, it is assumed that the temperature distributions in the cross-sections along the fuser roll axial direction are the same. This assumption reduces the Partial Differential Equation (PDE) problem from three-dimensional to two-dimensional.

Since the fuser roll is at the room when the operation starts, therefore, the initial condition is the following:

$$u(x,y,0) = T_\infty \qquad (x,y) \in \Omega \tag{2}$$

The geometry and important thermal properties of the system are listed in Table 1.

Table 1. Properties of the Pyrex transparent roll

Name	Value
Length L	0.2667m
Inner Radius R_i	0.011024m
Outer Radius R_o	0.012573m
Density of Pyrex ρ	2225kg/m^3
Specific Heat C_p	835J/kg·K
Thermal Conductivity k	1.4W/m·K
Ambient Temperature T_∞	297.44K
Lamp Wattage W	950Watts

The goal is to obtain an adequate glass fuser roll model which provides the transient temperature distribution and can be used in control system designs.

3 Approach and Results

The general idea of various modeling approaches to distributed systems is to approximate an infinite-dimensional system by a set of finite dimensional systems. Mathematically, this involves representing a distributed system (PDEs) with a set of Ordinary Differential Equations (ODEs). The finite element method (FEM) is a special type of weighted residual technique in which the basis and weight functions are non-zero on each small part of the spatial domain. FEM provides solutions everywhere within the boundary of the geometry and can be implemented on complicated geometries without too much difficulty (Ames 1992). A state-space model has been developed for an axi-symmetric heat conduction problem using the finite element approximation (Srinivasan et al 1994). In addition, due to the rapid development of the computer industry, computing power has been increasing constantly and computing costs have decreased drastically, which makes the computationally-intensive finite element method more and more practical for many applications.

This paper discusses the development of the FEM model for the transparent fuser roll using the Matlab PDE Toolbox. The toolbox provides a powerful and flexible environment for studying and solving PDEs in 2-D space and time by using the Finite Element Method.

Since solving the transparent fuser temperature distribution under the rolling fusing condition is a fairly complicated problem, it should be solved iteratively. Started with a set of much simplified boundary conditions, a stationary transparent roll heating up under the thermal radiation of the lamp is considered.

3.1 Stationary Roll Heating up

Both the inner and the outer boundaries of the roll are of Neumann type. On the inner boundary, insulation is assumed; on the outer boundary, the condition consists of heat influx due to the radiation heat transfer from the heating lamp to the black Teflon coating layer and convective heat transfer to the air (Incropera and DeWitt 1996). It is assumed that the black Teflon coating, which is about 0.0254mm thick, has no thermal mass and the radiant heat absorbed by the coating acts like heat flux to the outer boundary of the glass cylinder. The boundary conditions can, therefore, be summarized as the following:

$$\begin{cases} r = R_i, & \vec{n} \cdot (k\nabla u) = 0 \\ r = R_o, & \vec{n} \cdot (k\nabla u) = q - h(u - T_\infty) \end{cases} \quad (x,y) \in \partial\Omega \tag{3}$$

Where $\vec{n}$ is the normal unit vector of the geometry boundary, q is the heat flux due to the radiant heating, h is the convective heat transfer coefficient of the fuser roll surface, and $\partial\Omega$ is the boundary of the domain Ω.

The values of the coefficients in the boundary conditions are calculated based the conditions of the actual physical system. Since the heating lamp operates at 2500K as a blackbody, according to the Planck's law, the emissive power spectrum is shown in Figure 2. It indicates that 94% of total emitted energy is distributed within the 0.3μm and 3μm wavelength range. Notice that the transmissivity of the 1.55mm thick Pyrex glass within this wavelength range is approximately 92% (Touloukian and DeWitt 1972). Namely, 86.48% of the total power emitted by the lamp is transmitted to the black Teflon coating and absorbed by it. The rest 13.52% is absorbed by the glass roll, which will be modeled as an evenly distributed heat source Q within the glass, as in Equation (1).

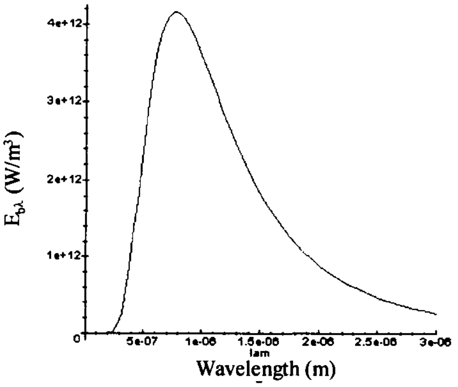


Figure 2. Emissive power spectrum of the lamp

Using a triangular mesh on domain Ω (see Figure 3) and at any time t ≥ 0, expand the solution to the equation (1) in the Finite Element basis, we have:

$$u(x,y,t) = \sum_i U_i(t) \cdot \phi_i(x,y) \qquad (x,y) \in \Omega \tag{4}$$

Substituting the expansion into equation (1), multiplying with a test function $\phi_j(x,y)$, integrating over Ω, and applying Green's formula and the boundary conditions (3) yields:

$$\begin{aligned} &\sum_i \iint_\Omega \rho C_p \frac{dU_i(t)}{dt} \phi_j \phi_i \, dxdy \\ &+ \sum_i \left(\iint_\Omega \nabla\phi_j \cdot (k\nabla\phi_i) + \int_{\partial\Omega} h\phi_j\phi_i \, ds \right) \cdot U_i \\ &= \int_\Omega Q\phi_j \, dxdy + \int_{\partial\Omega} q\phi_j \, ds \qquad \text{for } \forall j \end{aligned} \tag{5}$$

Using the method of lines to discretize equation (5) in 2-D space, we can assemble a system of ODEs in matrix form:

$$M\frac{dU}{dt} + KU = F \tag{6}$$

where M is the mass matrix, K is the stiffness matrix, and F is the force vector. While M is time-invariant, K and F has to be assembled repeatedly for each time step to incorporate the time-varying boundary conditions.

Given the initial value, the solution to equation (6) can be obtained at each time step by integration. It is noteworthy that the time derivatives of the nodal temperature can also be yielded easily. This is very important and convenient for adapting the FEM model into Simulink.

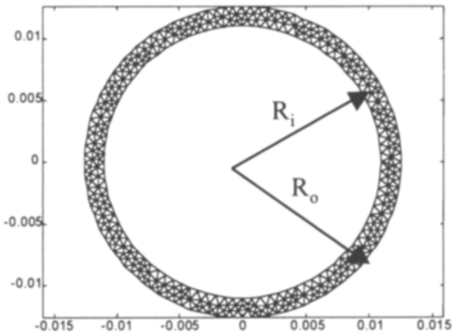


Figure 3. Cross-section of the roll (unit: meters)

The FEM model is ported into Simulink as an S-function because the flexibility of Simulink will enable various simulations to be carried out. In addition, future controller designs can be easily tested in the Simulink environment to evaluate their performance.

Figure 4 shows the Simulink diagram for the stationary transparent roll heating simulation. In the S-function "Stationary Roll", the boundary conditions of the PDE model change according to the input to the S-function. For example, when the heating lamp is off, there is no radiant heat flux going into the outer boundary. On the other hand, when the heating lamp is on, full strength heat flux is applied to the outer boundary. Since it is known that the heating lamp has a time constant of 0.3 seconds, a first order transfer function with the same time constant was added to model the dynamics of the heating lamp.

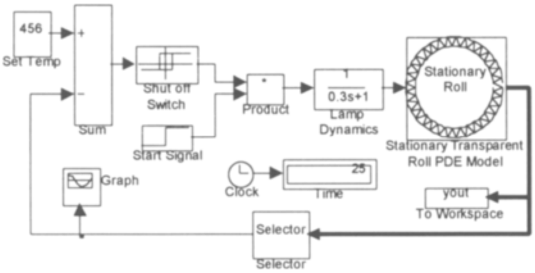


Figure 4. Stationary roll simulation diagram

The simulation process is as the following: the roll is at room temperature 297.44 K at the beginning. At Time = 1 second, the lamp is turned on and the roll starts to heat up. The PDE model solves the temperature distribution for the time step and outputs the temperature field. The "Selector" block is used to obtain the roll surface temperature from the output of the S-function. The surface temperature is then compared with the set value. If the surface temperature has not reached the desired point, the simulation will continue to the next time step. Otherwise, the "Shut off Switch" will be activated and the simulation will continue under the changed boundary conditions until the 25 seconds simulation time is reached. During the period after the lamp is shut off, the roll will cool down due to the nature convection.

The experiment was also carried out to verify the simulation results. The surface temperature was measured at a fixed point with an infrared pyrometer. Both the simulated and measured surface temperatures are plotted in Figure 5. It shows a good correlation between the two.

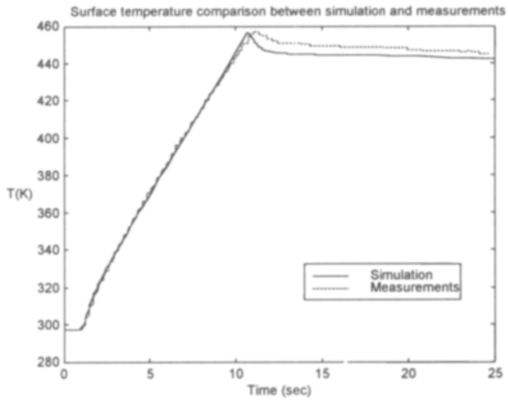


Figure 5. Surface temperature of the simulation result and the measured value

Since the stationary simulation result is satisfactory, a more complicated rolling fusing model can be built upon the stationary model.

3.2 *Simulation of the Fusing Conditions*

The simulation a rolling transparent fuser is achieved by changing the boundary conditions of the stationary model. Both the inner and the outer boundary conditions are still of Neumann type. On the inner boundary, insulation is assumed; on the outer boundary, depending on the time and location, the boundary conditions may be of two different forms: (1) heat influx due to the lamp thermal radiation minus the heat loss due to convection; (2) heat influx due to the lamp thermal radiation minus conduction heat transfer to the paper at the nip area. The boundary conditions can be summarized as the following:

$$\begin{cases} r = R_i, & \vec{n}\cdot(k\nabla u) = 0 \\ r = R_o, & \begin{cases} \text{Within the nip} \\ \vec{n}\cdot(k\nabla u) = q - h_{cp}(u - T_{paper}) \\ \text{elsewhere} \\ \vec{n}\cdot(k\nabla u) = q - h(u - T_\infty) \end{cases} \end{cases} \tag{7}$$

Where $h_{cp}=230 W/m^2K$ is the thermal contact conductance between paper and the transparent fuser roll. h is the convective heat transfer coefficient which is calculated based on the convection of a rotating cylinder (Mills 1992).

To simulate the rotating condition of the cylinder, the boundary conditions are set to revolve around the cylinder at the fuser roll angular velocity. In other words, the region where the roller contacts the paper is simulated to be traveling on the outer boundary of the geometry at a fixed angular speed, see Figure 6. The simulated sensor readings, by the same token, are calculated at two points traveling with the paper contact area on the outer boundary. This is achieved by programming the boundary condition inside the Matlab environment. The effect will be the same as having a roll rotating at -ω and everything else remaining fixed.

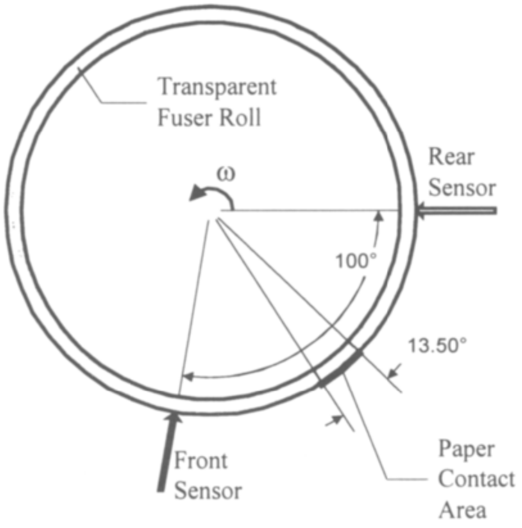

Transparent
Fuser Roll
ω
Rear
Sensor
100°
13.50°
Front
Sensor
Paper
Contact
Area

Figure 6. The rolling simulation scheme

Figure 7 shows the simulation diagram of the transparent roll during the fusing process. A simple fuzzy logic controller (FLC) is added to the system to regulate the fusing temperature. The controller uses both front and rear simulated sensor readings as inputs and calculates, heuristically, the output, which has a value within [0,1]. When the output value is 0, the total energy input to the lamp is 0 and the resulting effect is the lamp being turned off. On the other hand, when the output value of the fuzzy temperature controller is 1, the effect is the lamp being turned full on. The FLC output is passed to "Lamp Dynamics" block.

The simulated paper feeding is accomplished by using a pulse generator. It has an output value of either 1 or 0 representing with and without paper contact, respectively. Together with the output of the "Lamp Dynamics" block, the two values are passed to the transparent roll PDE model for every simulation time step to change the boundary values accordingly.

The simulation result is shown in Figure 8. The heating up process is almost identical to that of the stationary roll. At T=14 seconds, the paper feeding starts.

A better observation can be made with the zoomed-in plot in Figure 9. After the first 2~3 sheets, where transient responses are pronounced, the system reaches a pseudo steady state.

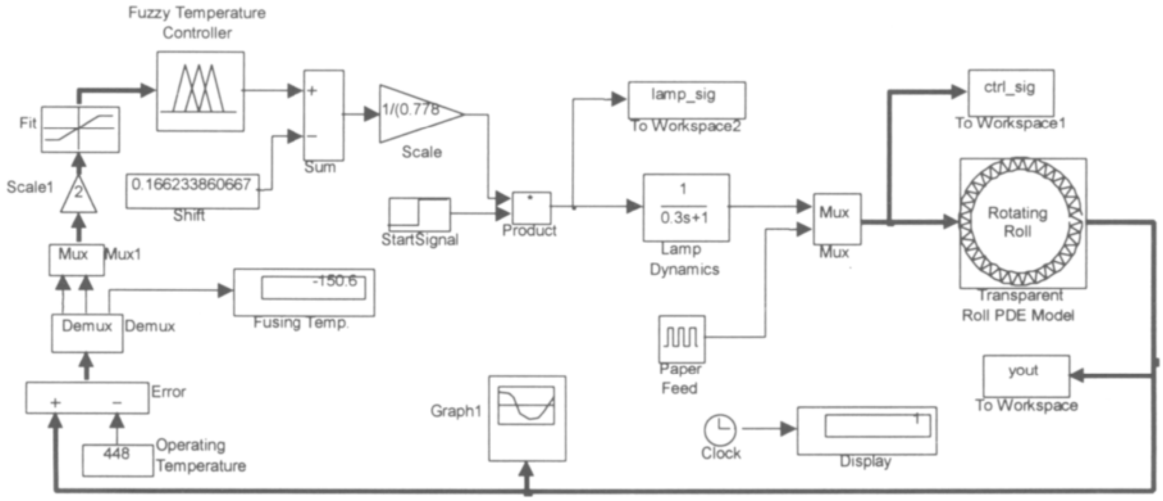

Fuzzy Temperature
Controller
Fit
+
−
Sum
1/(0.778
Scale
Scale1
2
0.166233860667
Shift
StartSignal
Product
Mux
Mux1
-150.6
Fusing Temp.
Demux
Demux
+
−
Error
448
Operating
Temperature
Graph1
lamp_sig
To Workspace2
1
0.3s+1
Lamp
Dynamics
Mux
Mux
Paper
Feed
Clock
1
Display
ctrl_sig
To Workspace1
Rotating
Roll
Transparent
Roll PDE Model
yout
To Workspace

Figure 7. Simulink diagram of the transparent roll during the fusing process

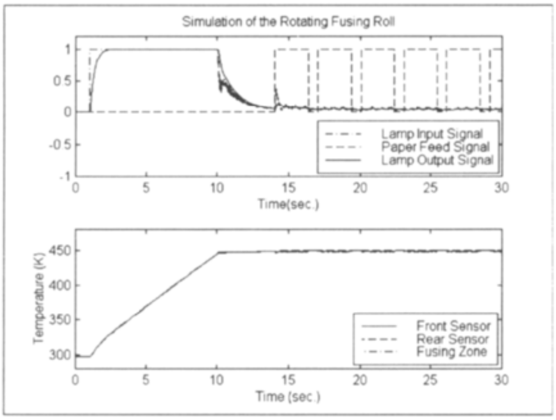


Figure 8. Results for the transparent roll with FLC

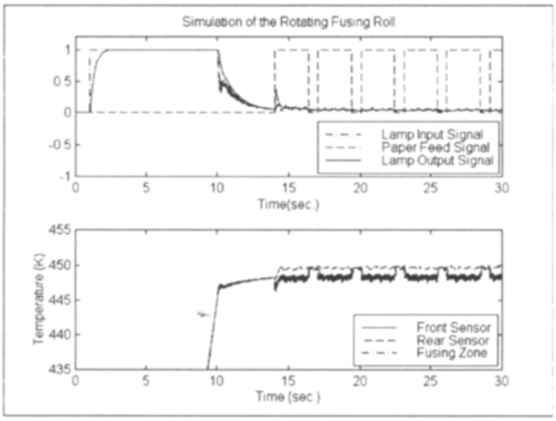


Figure 9. Zoomed-in plot of Figure 8

4 Conclusion and Future Work

An FEM model was developed and experimentally verified for the stationary transparent fuser roll. The result was satisfactory. An FEM model was also developed for the rolling fusing transparent fuser roll. A simple fuzzy logic controller was implemented with the FEM model. Simulation was carried out and reasonable result was achieved.

The scientific contributions of this work are the development and verification of an FEM-based model for a distributed, non-steady state, thermal system for the purposes of control design. Even though the research is based on the transparent xerographic fuser, the methodology used in the process is applicable to a broader class of applications.

The future work includes designing and simulating more sophisticated controllers using state-space techniques. Comparisons will be made between the simulated results of various control designs. Implementations will be carried out to validate the control designs.

Reference

Ames, W. F., 1992. *Numerical Methods for Partial Differential Equations, 3^{rd} ed.* Boston: Academic Press, Inc..

Srinivasan, A., Batur, C., and Rosenthal, B., 1994. Control of thermal system through finite element based state space model. *Proceedings of the American Control Conference, Baltimore, Maryland, USA, June 1994.*

Incropera, F. P. and D. P. DeWitt, 1996. *Introduction to Heat Transfer, 3^{rd} ed.* New York: John Wiley & Sons.

Touloukian, Y. S. and D. P. DeWitt, 1972. *Thermal Radiative Properties: Nonmetallic Solids, vol. 8 of Thermophysical Properties of Matter.* New York: Plenum Press.

Mohr, J. W. , et al, 1997. Thermal Contact Conductance of a Paper/Elastomer Interface. *Journal of Heat Tranfer*, 119(5), 363-366.

Mills, A. F., 1992. *Heat Transfer.* Homewood: Irwin

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Experiment to evaluate the feasibility of the Smart Disc concept

J. Holterman, T.J.A. de Vries, M.P. Koster

Cornelis J. Drebbel Institute for Systems Engineering
Department of Electrical Engineering, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
Phone: +31-53 489 27 88, Fax: +31-53 489 22 23
E-mail: mechatronics@el.utwente.nl, www-address: http://www.rt.el.utwente.nl/mechatronics

Keywords: mechatronics, vibration control, flexible frames, piezo actuation, piezo sensing

Performance of contemporary high-tech equipment is more and more determined by accuracy. Positional deviations due to high frequent vibrations in the floor can be suppressed by mounting the equipment on a frame that acts as a low-pass filter, but low frequent deviations and deviations due to motions inside the machine can only be reduced by increasing the machine frame stiffness. When only passive elements are used during design, large stiffness can only be attained with rigid bodies, which often exhibit bad dynamic behaviour.

In case a machine should be both accurate and fast, a design with active elements should be considered. The Smart Disc is an example of such an active element. It integrates a piezoelectric sensor, a piezoelectric actuator and control electronics, such that it expands when it experiences a compressing force, and vice versa. Insertion of Smart Discs at discrete locations in a machine might compensate for the lack of frame stiffness.

To evaluate the feasibility of the Smart Disc concept an experimental set-up has been designed. Though only one prototype Smart Disc is utilised, the set-up resembles a large class of machines suffering from deviations caused by the non-infinite frame stiffness and non-filtered vibrations in the floor. Experiments with this set-up show a reduction of the positional error of interest in the order of 20 dB in a frequency range of 0.1 to 10 Hz.

1 INTRODUCTION

A machine frame has two main functions. First, it takes care of the relative positioning of machine parts, and second, it provides supporting forces in order to carry the load of machine parts. As the desired accuracy of operation of machines increases, these two functions become hard to combine. The deflection of the machine frame due to load carrying will be of the same order of magnitude as the allowed positioning errors.

When only passive elements are used during design, the above-mentioned problem can only be solved with large stiffness. Large stiffness however can only be attained with rigid bodies, which often exhibit bad dynamic behaviour. Therefore 'designing', with passive elements only, is searching for a compromise between accuracy and speed. Adequate performance may then not always be attainable for acceptable costs. In that case the use of active elements can help to increase performance or reduce costs.

This paper describes the design and the results of an experiment that may prove the feasibility of an 'active damping' concept with so-called Smart Discs. A Smart Disc is thought of as an integrated device that expands when it experiences a compressing force, and shrinks when it experiences a stretching force. This 'recalcitrant' behaviour can be realised by integration of a piezoelectric sensor, a piezoelectric actuator and control electronics. Insertion of Smart Discs at discrete locations in a machine might compensate for the lack of stiffness of the frame, i.e., reduce positional deviations at certain other locations of interest in the machine.

The number of applications one can imagine for a Smart Disc-like device is quite large. Nevertheless in this paper only a special class of applications is considered. The main function of the equipment this class covers, is accurate positioning of the top of a non-rigid frame with respect to the base of the frame, as will be explained in the next section. The subsequent sections describe the design of a suitable yet simple experimental set-up, as well as the results of experiments with this set-up.

2 PROBLEM DESCRIPTION

To evaluate the feasibility of the Smart Disc concept, this paper concentrates on a special class of applications. This class covers accurate high-tech equipment like electron microscopes, wafersteppers and telescopes. The main function of the frame of these machines is accurate positioning of the top of the frame with respect to the base of the frame.

Usually accurate equipment like this is mounted on a resiliently supported frame that filters the vibrations from the 'fixed' world. As the equipment is most sensitive for high frequent disturbances, this mass-spring configuration, acting as a low-pass filter, is rather appropriate. However, due to the limited filter bandwidth, the top of the frame will still suffer from some low frequent deflection (Fig. 1a).

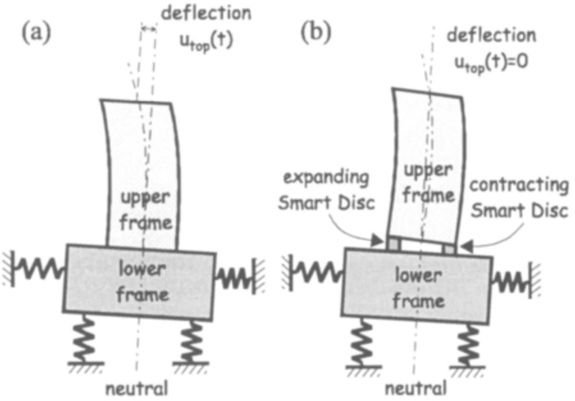


Figure 1. a) Resiliently supported frame.
b) Smart Discs compensate for deviation.

Until now the low frequent disturbances, for instance lower than 10 Hz, hardly affected the performance of the equipment. But with increasing accuracy demands these disturbances can no longer be ignored. One should either reduce the disturbances, by decreasing the cut-off frequency of the low-pass filter, or reduce the effect of the disturbances on the equipment, by increasing the frame rigidity. Both solutions are based on the optimal design of passive elements and may not be satisfying. In that case the use of active elements, Smart Discs for instance, should be considered.

2.1 Smart Disc concept

For a good understanding of the Smart Disc concept, it is important to note that the deviation relevant for accurate operation is not the movement of the complete machine (heavy lower frame together with flexible upper frame), but the relative movement between the upper frame and the lower frame, as indicated in Fig. 1. The proposed solution with active elements is illustrated in Fig. 1b. In this planar perception the upper frame is connected to the lower frame by means of two Smart Discs. In a three-dimensional configuration, in which a frame can deflect in more than one direction, the frame should be supported by at least three Smart Discs.

The Smart Disc concept in essence boils down to the following. The *actuators* of the Smart Discs can compensate for the deformations in the upper frame through rotating the base of the frame. Appropriate expansion respectively contraction of the Smart Discs can accomplish the desired amount of rotation.

When one, for instance for economic reasons, insists on a spatially integrated single device, the deformation has to be estimated with the use of information from the *sensors* of the Smart Discs. If only the lowest mode of vibration is considered, the deflection at the top can be estimated from the load between the upper and the lower frame. This estimation is only possible if each Smart Disc is equipped with knowledge of the actual set-up it is utilised in, such that the *control electronics* can calculate the necessary amount of expansion or contraction.

An envisioned realisation of the integration of the three main elements of the Smart Disc (piezoelectric sensor, piezoelectric actuator, control electronics) is depicted in Fig. 2.

3 EXPERIMENT DESIGN

The Smart Disc concept for the class of applications considered in this paper is based on a rotation of the base of the upper frame in order to reduce the positional error at the top of the frame. To experimentally evaluate the feasibility of this concept, it suffices to design a set-up in which only one prototype Smart Disc is used, as will be made clear in this section.

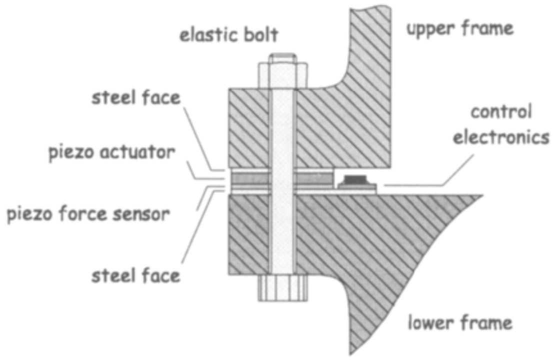


Figure 2. Envisioned realisation of the Smart Disc.

3.1 Positional error caused by force on top

The majority of the deflection u_{top} in Fig. 1 is due to the lowest mode of vibration of the frame. An approximate model to analyse this is given by a discrete mass and a discrete stiffness, both at the top of a rigid massless frame (Fig. 3). Herewith only the lowest eigenfrequency of the frame is captured.

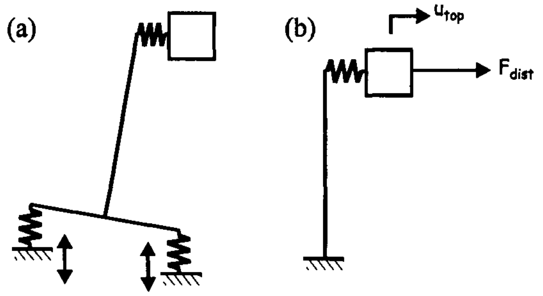


Figure 3. Approximate model of a non-rigid frame, the deformation of which is caused by
a) vibrations in the floor,
b) an external force on top.

From the approximate model it can be seen that when the floor suffers from vibrations, the rigid massless frame will rotate or translate. As a consequence the spring at the top of the frame will expand or contract, due to the inertia of the mass it is connected to, and in turn this mass will experience a resulting force (Fig. 3a). Note that in this model the amount of expansion or compression of the spring represents the positional deviation of interest.

From the above considerations it follows that the deviation at the top of the non-rigid frame can be thought of as the effect of a resulting force on the top of the frame. Vibrations in the 'fixed' world thus cause the same kind of positional deviations as an external disturbing force on the top of a non-rigid frame in a non-moving world. Therefore, to evaluate the feasibility of the Smart Disc concept, it suffices to consider a frame in a non-moving world upon which a disturbing force acts at the top (Fig. 3b).

3.2 Fixed rotation point

The deviation at the top of the upper frame can be reduced or even completely compensated for by rotation of the base of the upper frame as illustrated in Fig. 1b. The rotation of the base can be accomplished by appropriate expansion respectively compression of the two Smart Discs.

In order to keep the same average height, independent of the angle of rotation, the thickness variation of both Smart Discs must be the same. In that case a fictitious rotation point halfway the two Smart Discs can be defined (Fig. 4a). Therefore, to evaluate the feasibility of the Smart Disc concept, it suffices to implement a physical point of rotation (the hinge point) and only one prototype Smart Disc in the experimental set-up instead of two (Fig. 4b).

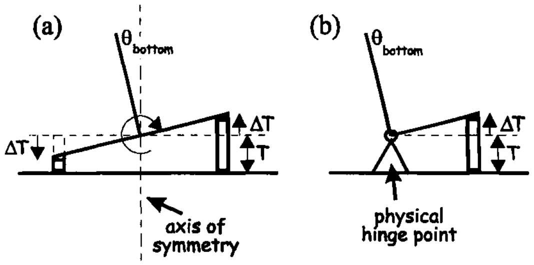


Figure 4. a) Symmetry in the set-up.
b) One Smart Disc instead of two.

3.3 Experimental set-up

With the previous considerations in mind, the experimental set-up has been designed (Fig. 5). The machine frame is modelled by a simple beam of steel with finite stiffness (dimensions: 200x50x10 mm^3). The hinge point in the lower corner of the frame (a so-called 'gatscharnier'; Koster, 1998) allows only for a small, but sufficiently large amount of rotation.

This implementation of the hinge point indeed constitutes a fixed rotation point, as it is hysteresis-free.

The disturbing force is generated by an electromagnet, which is mounted on a separate beam. By the choice of an electromagnet the disturbing force can easily be controlled.

To be able to evaluate the effect of the Smart Disc action, the deflection of the top of the frame should be measured. For this purpose a linear variable differential transformer (LVDT) is mounted on a separate beam. Note that this additional measurement is not used in the control action of the Smart Disc actuator.

In the experimental set-up use is made of a prototype Smart Disc in which a personal computer takes the place of the control electronics. The prototype Smart Disc is treated in the next section.

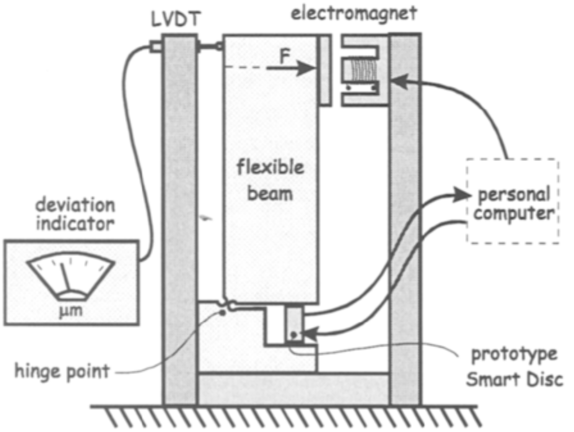


Figure 5. Schematic view of the experimental set-up.

4 SMART DISC

As the prototype Smart Disc used in the experimental set-up has not yet been equipped with control electronics, the Smart Disc only integrates an actuator and a sensor. For both the actuator and the sensor use is made of piezoelectric material known as CMA (ceramic multilayer actuator; refer e.g. Miu, 1993), referred to in this paper as piezo mono-stacks.

The prototype Smart Disc consists of two compound piezo stacks (height: 8.4 mm), which in turn both consist of three piezo mono-stacks (dimensions: 7×7×2 mm^3; see Fig. 6). Two of the mono-stacks are used for actuation and the other is used for sensing. Though in theory it is possible to use only one piezo mono-stack as both a sensor and an actuator, in practice the voltage to be measured (~ 1.5 V) would be disturbed too much by the actuating voltage (~ 60 V maximum).

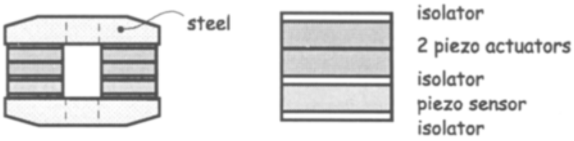


Figure 6. Prototype Smart Disc (left), built up of two compound piezo stacks (right).

From the dimensions of the piezoelectric layers the piezo mono-stacks consist of (thickness: 20 µm), it can be deduced that the lowest possible resonance frequency of the piezo stacks is above 50 kHz and thus far above the bandwidth relevant to the experiment. Therefore it is reasonable to neglect the dynamic behaviour of a piezo and only to use a quasi-static model (Geraeds, 1996b). A single piezo mono-stack with d_{33}-coupling (i.e., the direction in which the piezo extends is parallel with the poling axis) can then be described with the following matrix equation:

$$\begin{bmatrix} \Delta t_h \\ \Delta Q \end{bmatrix} = \begin{bmatrix} C_{mech} & d \\ d & C_{elec} \end{bmatrix} \cdot \begin{bmatrix} F_3 \\ U \end{bmatrix}, \tag{1}$$

with

Δt_h : thickness variation of the piezo [m],
ΔQ : excess charge on the piezo [C],
F_3 : force applied to the piezo [N],
U : voltage across the piezo [V],

and

C_{mech} : mechanical compliance of the piezo [m/N],
C_{elec} : electrical capacity of the piezo [C/V],
d : piezoelectric charge constant [C/N] ≡ [m/V].

From eq. (1) it can be seen that, due to the piezoelectric coupling phenomenon (d), a piezo can be used both as force sensor and as position actuator. The maximum thickness variation of the single mono-stacks used in the prototype Smart Disc, when the maximum allowed voltage (60 V) was applied, was 2.42 µm.

4.1 Force sensor

When a piezo is used as a force sensor, two designs are possible. The first design is based on zero voltage across, i.e. short-circuiting of the two conducting layers of the piezo. The current through the short circuit then is a measure for the time derivative of the applied force. The second design is based on zero current between the two conducting layers. In this case the voltage across the piezo is a measure for the applied force. As in the latter design the measured signal needs no integration, the piezo force sensor is implemented as a voltage generator.

In the voltage generator configuration, the resistance between the conducting layers is near infinity. As a consequence the initial charge on the piezo is undefined and will drift as a result of external electric disturbances. To prevent malfunctioning of the sensor, the initial charge is forced to zero, by connecting a known shunt resistance parallel to the piezo.

Furthermore, instead of an ideal voltage meter, a non-inverting amplifier with an input resistance near infinity is used. The implementation of the force sensor is shown in Fig. 7. Here the piezo is modelled as a voltage source in series with a capacitor. The transfer function from voltage across the piezo to measured voltage now can be given by

$$U_{out}(j\omega) = K\frac{j\omega}{j\omega+\omega_0}\cdot U_{piezo}(j\omega), \tag{2}$$

in which we recognise a high-pass filter with gain

$$K = 1+\tfrac{R_2}{R_1}. \tag{3}$$

An important consequence of this implementation is that the force sensor does not operate for frequencies below the cut-off frequency

$$\omega_0 = \left(C_{elec}R_{shunt}\right)^{-1}. \tag{4}$$

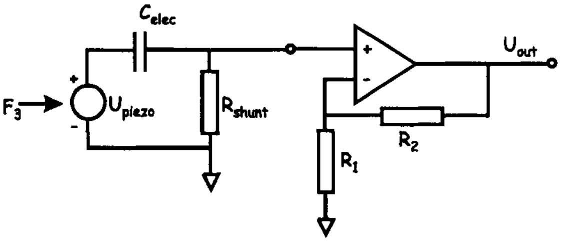


Figure 7. Implementation of the force sensor.

In the experiments use was made of a force sensor implemented according to Fig. 7 with cut-off frequency 0.13 rad/s.

4.2 Position actuator

When a piezo is used as a positional actuator, again two designs are possible. Usually a design based on voltage control is used because of its simple implementation. The thickness variation of a mechanically unloaded piezo then is linearly dependent to the applied voltage. An important disadvantage of voltage control, 'hysteresis', is almost completely solved when using the other possible design, based on charge control (Main et al., 1995). Here the thickness variation of the mechanically unloaded piezo is linearly dependent to the applied charge. Though the amplifier needed for charge control is more complex, in the experimental set-up use is made of a charge controlled position actuator.

4.3 Controller design

When the dynamics of the experimental set-up are neglected, the set-up can be represented as a multivariable model with only simple gain factors.

$$\begin{bmatrix} u_{top} \\ U_{piezo} \end{bmatrix} = \begin{bmatrix} k_1 & k_2 \\ k_3 & k_4 \end{bmatrix} \cdot \begin{bmatrix} F_{ext} \\ U_{act} \end{bmatrix} \tag{5}$$

Here we recognise

u_{top} : deviation to be compensated for [m],
U_{piezo} : voltage across the force sensor piezo [V],
F_{ext} : external force at the top of the frame [N],
U_{act} : voltage across the position actuator piezo [V],

and

k_1 : set-up overall compliance, $0.467\cdot10^{-6}$ [m/N],
k_2 : piezo actuator constant, $-6.22\cdot10^{-6}$ [m/V],
k_3 : piezo sensor constant, 0.15 [V/N],
k_4 : piezo cross-talk constant, 1.95 [-].

The values for the four gain factors have been determined experimentally.

With the use of the simple static model a rather straightforward model-based controller has been designed. This has been motivated by the fact that the emphasis in this experiment was not on controller design but on the evaluation of the feasibility of the Smart Disc concept.

5 RESULTS

With the set-up designed as described in section 3 and 4, two experiments have been performed. The purpose of the first experiment was to determine the frequency domain damping characteristics of the prototype Smart Disc. A sinusoid disturbance force, with amplitude 25 N was applied to the set-up, in a frequency range from 0.01 to 7 Hz. In case of no active damping this caused a sinusoidal deviation with amplitude 12.5 μm. The damping the utilisation of the prototype Smart Disc yielded is depicted in Fig. 8.

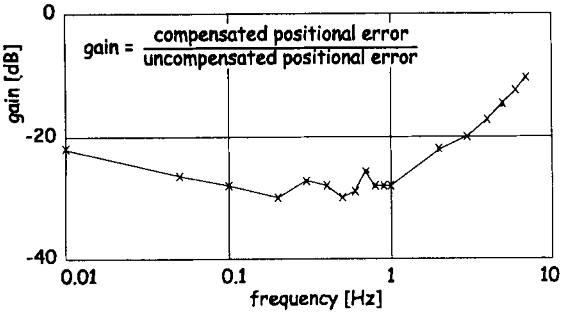


Figure 8. Frequency response of the active damping.

In the second experiment the response of the set-up was evaluated in the time-domain. Again the set-up with active damping was compared to the uncompensated set-up. The applied force was a 2 Hz block-shaped wave, resembling deformations due to non-filtered motions inside the machine, thus cursed with high frequent components. From the results (Fig. 9) it can be seen that the Smart Disc indeed is able to compensate for low frequent deviations. However, high frequent disturbances are hardly suppressed.

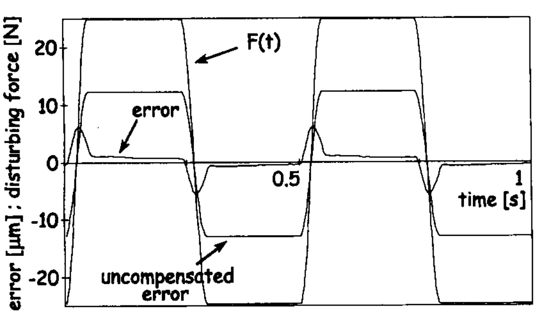


Figure 9. Response to a block wave shaped force.

6 CONCLUSION

To study the principle functionality of a Smart Disc, it suffices to have a set-up in which a disturbance force acts on the top of a non-rigid beam with a hinge point in one lower corner and a Smart Disc supporting the other lower corner (Fig. 5).

Experiments with the set-up show that the Smart Disc can compensate for a positional deviation at the top of the upper frame. Within the given disturbance bandwidth of 0.1 Hz up to 10 Hz the Smart Disc can realise an attenuation of approximately 10 to 30 dB. The remaining disturbances mainly originate from a higher harmonic cross-talk between the steered actuator voltage and the measured sensor voltage.

These initial results suggest that the concept may be feasible in reality.

ACKNOWLEDGEMENT

The authors wish to thank Pascal Geraeds; the material for this paper has mainly been taken from his M.Sc. thesis (Geraeds, 1996a).

REFERENCES

Geraeds, P.M.J., (1996a), *Active damping of low frequent vibrations in non-infinite stiff frames with a Smart Disc (an experiment design)*, M.Sc. thesis, Report no 029R96, Control Laboratory, Electrical Engineering Department, University of Twente, Enschede, The Netherlands.

Geraeds, P.M.J., (1996b), *Overview of piezoelectric ceramics*, Report no 034R96, Control Laboratory, Electrical Engineering Department, University of Twente, Enschede, The Netherlands.

Koster, M.P., (1998), *Constructieprincipes voor het nauwkeurig bewegen en positioneren* (in Dutch), Twente University Press, Enschede, The Netherlands.

Main, J.A., E. Garcia, D.V. Newton, (1995), Precision Position Control of Piezoelectric Actuators Using Charge Feedback, *Journal of Guidance, Control and Dynamics*, Vol. 18, No. 5, September-October, pp. 1068-1073.

Miu, D.K., (1993), *Mechatronics, Electromechanics, Electromechanics and Contromechanics*, Department of Mechanical, Aerospace and Nuclear Engineering, University of California, Los Angeles, Springer Verlag New York, USA.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Modelling and Motion Control of an Autonomous Underwater Robot using a Fuzzy and Fuzzy-neural Approach

I. S. Akkizidis and G. N. Roberts

Mechatronics Research Centre, University of Wales College, Newport,
Allt-yr-yn Campus, PO Box 180, Newport, South Wales, NP9 5XR, UK

The design of an autopilot for the control of an Autonomous Underwater Vehicle (AUV) is of interest both from the point of view of motion stabilisation as well as manoeuvring performance. Such vehicles are commonly classified as being highly non-linear uncertain systems and are consequently difficult to control effectively using model-based control methods. The modelling task of these vehicles which consists of three-dimensional equations of motion for its hydrodynamical shape is very complicated. The paper describes a modelling, mission and motion control system for an AUV based on fuzzy and fuzzy neural techniques. The idea for the fuzzy modelling is an on-line supervisory scheduling system, which chooses the linear function that describes the system. Fuzzy mission control and fuzzy neuro motion control strategies, with the ability to adapt membership functions, extract new rules and forget unused rules, are proposed.

1. INTRODUCTION

Manoeuvring control of underwater vehicles is a demanding task. This difficulty stems from the fact that Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) may be classified as uncertain systems possessing highly non-linear dynamics. Therefore developing the model of the AUV dynamics has to be attempted before controller design may be considered. AUVs and ROVs are highly non-linear, multivariable dynamic process which means that controller designs must be able to deal with the non-linearities encountered.

This study described herein is based on a low-cost ROV named GARBI developed at the University of Barcelona and the University of Girona. The vessel, which is illustrated in figure 1, is used for underwater mission operations such as observations and inspections [1,2,3]. GARBI is linked to a surface ship (or with other operating platform) by an umbilical cable carrying power and providing a communication link. This umbilical link imposes limitations to the ROV operations, such as depth limits and danger of cable snagging.

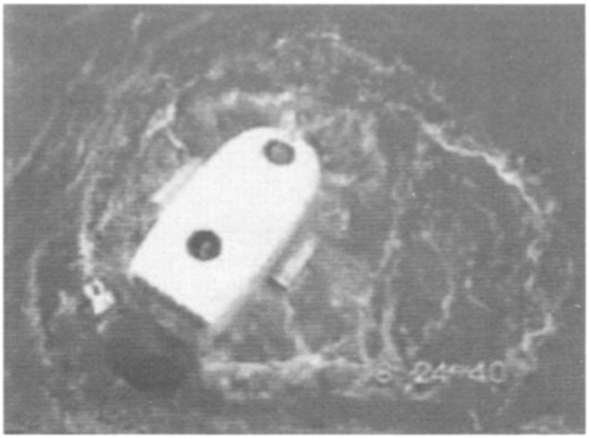


Figure 1. Photo of GARBI underwater robot

By definition AUVs are vessels which have sufficient on-board power and intelligent ability to move purposefully without human intervention in environments that have not been specifically engineered for it. It is clear therefore that AUVs bring a number of advantages: they have no umbilical cable to limit range, or to become entangled in a surrounding structure. They can also undertake missions that would be impractical or impossible with an ROV, such as long range observations and collection of oceanographic data and under-ice surveying.

The motivation for this work is to extend GARBI from operating as an ROV to operating as an AUV. Fuzzy Logic and Neuro Fuzzy techniques are propose for modelling, mission and motion control.

1.1. Comparison between Artificial Neural Network and Fuzzy Logic

A detailed presentation of Fuzzy Systems (FS) and Artificial Neural Networks (ANN) theory is beyond the scope of this paper. A brief comparative study between FS and ANNs related to their operations in the context of knowledge acquisition, uncertainty, reasoning and adaptation is presented in table 1.

1.2. Fuzzy Neural Control systems (FNC)

It can be seen from table 1 that the advantages of the fuzzy approach are mainly the disadvantages of the ANN approach, and vice versa. So the idea is naturally to combine neural networks and fuzzy systems to overcome their disadvantages, but to retain their advantages. The integration of these two fields has given birth to Fuzzy-Neural Systems (FNS). From ANN, the powerful learning capabilities enable these systems either to automatically learn the fuzzy decision rules or to learn from a set of plant measurements, whereas the fuzzy presentation enables designers to extract the learnt information from the expert in a form easily understandable.

2. MODELLING OF AUV

2.1. Fuzzy Modelling

Fuzzy Modelling is the method of describing the characteristics of a system using fuzzy rules. The technique can express complex non-linear dynamic systems by linguistic *if*-then rules [4].

A typical Takagi-Sugeno fuzzy model has the form:

$$R_s^i \quad if \quad s = LS^i \quad then \quad x_i = f_i(x,u) \tag{1}$$

where s is the operating point vector $s = \{s_1, s_2, \ldots, s_{n_s}\}$. The vector consists, in general, of state, input and output variables (the n_S is its dimension). LS^i is the i-th fuzzy state vector equal to:

$$LS^i = (LS_1, LS_2, \ldots, LS_{n_s})^T \tag{2}$$

where LS_{n_s} is the fuzzy values of s with appropriate Membership Function (MF).

The *then*-part of this fuzzy rule defines a linear autonomous open loop model representing the system dynamics within the fuzzy region LS^i specified in the *if*-part of the same fuzzy rule. This model is of the form $x_i = f_i(x,u)$, where f_i is a linear function normally obtained via an identification procedure.

The reason for adopting the fuzzy model described above is that most systems (or plants) are non-linear and therefore cannot be described by a

Table 1. A comparative study between fuzzy systems and artificial neural networks

Skills		**Fuzzy Systems**	**Artificial Neural Networks**
Knowledge acquisition	Inputs	Human experts	Sample sets
	Tools	Interaction	Algorithms
Uncertainty	Information	Quantitative and qualitative	Quantitative
	Cognition	Decision making	Perception
Reasoning	Mechanism	Heuristic search	Parallel computations
	Speed	Low	High
Adaptation	Fault-tolerance	Low	Very high
	Learning	Induction	Adjusting synaptic weights
Natural language	Implementation	Explicit	Implicit
	Flexibility	High	Low

single linear model. Instead of constructing complicated non-linear models based on physical laws, an alternative approach can be used, namely the construction of a collection of linear models. In this case, each (local) linear model approximates the original non-linear system around different operating points and a *supervisory scheduling system* determines which local linear model suits the particular operating conditions [5].

The supervisory scheduling system uses a discriminant function:

$$w = f_d(s) \tag{3}$$

This function defines a weight vector $w = \{w_1, w_2, \ldots, w_k\}$, with $w_i \in [0,1]$, for each particular value of the operating point vector. This definition can be applied using *min* or *max* operation. For example, for the particular crisp value $s = \{s_1^*, s_2^*, \ldots, s_{n_s}^*\}$ of the operating point vector the weight vector will be:

$$w = \min(\mu_{LS_1}(s_1^*), \mu_{LS_2}(s_2^*), \ldots, \mu_{LS_{n_s}}(s_{n_s}^*)) \tag{4}$$

where $\mu_{LS_{n_s}}$ is the degree of membership of the crisp value $s_{n_s}^*$ of s_{n_s}.

The overall output X of the composite model is calculated as the weighted mean of the outputs x_i of the local linear open loop models, given as:

$$X = \frac{\sum_{i=1}^{k} w_i \cdot x_i}{\sum_{i=1}^{k} w_i} \tag{5}$$

where k is the number of local open loop linear models.

2.2. Linear modelling study of GARBI

GARBI is a non-linear, multivariable dynamic system. In order to design a controller for GARBI or to undertake simulation studies a dynamic model is needed. Mathematical models, which describe these non-linear dynamics, are very complex as well as limited in terms of only being able to represent the physical laws of the system. An identification method based on experiments is proposed. Using linear control theory (local) models can be investigated. Step or frequency response methods can be applied to identify the transfer functions which describe the relationships between inputs and outputs of the system, in the *s*- or *z*- domain.

The equivalent difference equations of the transfer functions can be written in the form:

$$x(t) = -\sum_{i=1}^{n_a} a_i x(t-i) + \sum_{j=0}^{n_b} b_j u(t-j-1) \tag{6}$$

where the degrees n_a, n_b of the polynomials are given by the order of the backward shift operation, n_a = number of poles of G(s) and $n_b \le n_a$.

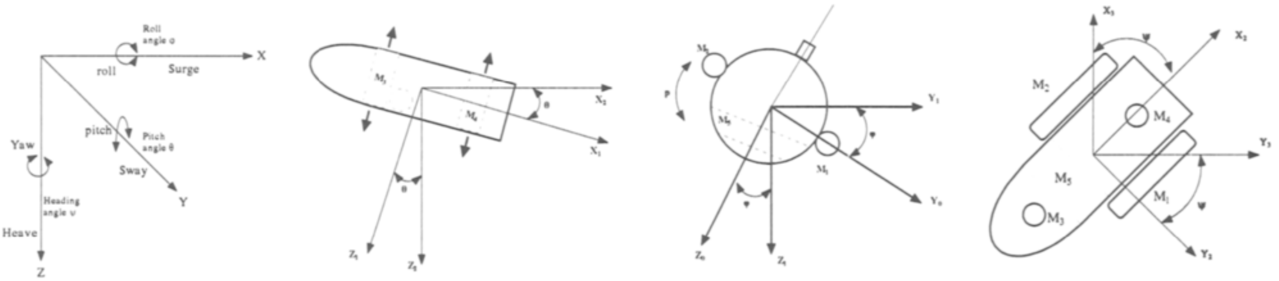
However, using fuzzy model theory the GARBI non-linear system can be described as in section 2.3.

2.3. Fuzzy Modelling of GARBI

The study of the motion of marine vehicle involves six *Degrees Of Freedom* (DOF) (Fig. 2a), since six independent co-ordinates are necessary to determine the position and orientation of a rigid body. The first three co-ordinates (Surge, Sway and Heave) and their time derivatives correspond to the position and translational motion along the *x-, y-, and z-axes*. The last three co-ordinates (roll, pitch and heading angle) and their time derivatives are used to describe orientation and rotational motion.

In order to drive the GARBI on the (x,y), (x,z), (y,z) surface, four different speeds (Zero, Low, Medium and High) of its five propellers M_1 to M_5 (Fig. 2b,c,d) have to be applied. These are the inputs of the system's model. The output however, depends on the six DOF of the vessel. Therefore, the rotation about the *z-axes* may be responsible for roll & yaw, the *y-axes* for pitch, surge & heave, and the *x-axes* for sway, roll & yaw. The functions that characterise the relation between this Multi-Input Multi-Output (MIMO) system represent each (local) linear model of the vessel.

The fuzzy variables with speed in the region *Zero, Low, Medium* and *High* can be used to fuzzify the input of the system's model. It is considered that the shape of the Membership Function within these regions is not so critical. The shapes can be chosen after experimental tests.



(a) degrees of freedom (b) side view (c) horizontal reference (d) top view

Figure 2. Illustration of the GARBI, showing its degrees of freedom and control

The fuzzy model is given as a set of fuzzy rules with each fuzzy rule R_s^i being of the form as shown in (eq.1). Specifically in the GARBI's modelling the rules will have a form such as:

R_s^i : **if** *speed Mi* is High **and** *speed Mj* is Low **then** x=f(x,y)

The *then*-part of the above fuzzy rule defines a linear autonomous open loop model representing the system dynamics in terms of the six DOF of GARBI, within the fuzzy regions (Zero, Low, Medium and High) specified in the *if*-part of the same fuzzy rule.

Therefore, during the fuzzification procedure the fuzzy inputs are defined. Using Sugeno type fuzzy rules in the knowledge base, the local open loop linear models are determined (equation. 1). The sum of these models, (equation 5), having different weight, will describe the non-linear GARBI system for any variation in the input and/or output.

3. CONTROLLING THE AUV

3.1. Fuzzy-Neuro Control

Fuzzy control is a successful application of fuzzy theory [6,7] and is used in many applications [8]. Fuzzy Logic Controller (FLC) is a knowledge-based controller, which uses fuzzy logic for knowledge representation and inference [9]. One feature of FLC is that expert knowledge is represented by a compact set of fuzzy *if-then* rules consisting of membership functions in the *if*-part and functions agreeing with the controller in the *then*-part. Therefore, the design of the parameters in FLCs is obtained from the expert. However, FLCs cannot reflect a non-linear system dynamics fully. In addition, the design processes depend on trial-and-error methods or some heuristic algorithm [10]. Moreover, an expert who knows the characteristics of the system may also be needed for setting up the initial rules.

If the actual output of the controller differs from the desired behaviour, either an unsuitable choice of membership functions or missing fuzzy rules is possible. The desired output can be estimated either under simulation or from experimental results.

Multi-layer neural networks with learning algorithms, have been used for *extracting fuzzy rules* or *tuning (adapting) the membership functions* in a linguistic variable based fuzzy control environment.

Here it is proposed to first try and solve the problem using adaptation of membership function technique. If the solution is not satisfactory, a technique for extracting new fuzzy rules is used.

3.1.1. Adaptation of Membership Functions

The adaptation of membership functions is a reverse mechanism that is deduced from the forwarding inference mechanism. The approach comes from ANNs, where the computation of the control values are obtained from the measured input values, which are feeding a feedforward procedure similar to that found in layered neural nets. If the actual output is not able to drive the controlled system to a desired state, an error has to be propagated back through the architecture, changing the parameters, taking into account the feed forward propagation of inputs.

If it is not possible to determine an optimal control action for a given state, the error of the produced output cannot be calculated directly. Therefore, a reinforcement or "non-supervised" learning algorithm is used to evaluate the error of the fuzzy controller corresponding to the given input [11]. However, if the desired control output is known, then a "supervised" learning algorithm is used and the network is trained corresponding to that output. Previous work in this field is reported in [12,13,14]. In each case it is the range of membership functions, which is re-adjusted.

3.1.2. Extraction of new fuzzy rules

In some cases, the expert cannot predict all the rules to construct the base-knowledge of the fuzzy controller. Also it is possible that some of the rules may not be the most suitable. Algorithms to extract new fuzzy rules can be considered as one solution to this problem [15].

An alternative approach, which is used in this study, is to extract new fuzzy rules as above but also to correct existing rules and to forget unused or incorrect rules.

There are several ways to investigate why the fuzzy controller may fail because of improper rules, such as the error of the controller is too high or the controller reaction is too slow. The method is first to try and solve the problem by adding or taking out linguistic variables of the antecedent part of the fuzzy rules. If the improvement is not so good then new rules are created. If a rule is not used, then it is removed.

3.2. Control study for GARBI

The motion control system for GARBI shown in figure 3, involves two control loops: one for mission control and one for navigation. The parameters of the mission target such as depth and field of action are specified using fuzzy logic control techniques. The navigation of GARBI, which is affected by currents, obstacles and turbulence (regarded, as disturbances) is developed using Fuzzy-Neuro control technique as discussed above. Therefore the motion controller using the five propellers (Fig. 2 b,c,d) can reach the mission target and also be robust against disturbances. Typical Mamdany and Sugeno fuzzy model rules are used to construct the knowledge base, including the mission and the control data respectively.

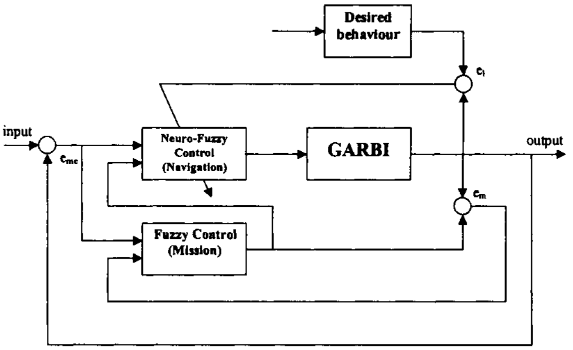


Figure 3. Motion Control System

The fuzzy rules have the form:

For the mission control:

$$R^{i}_{cm} \quad if \quad cm_a = LA_i, \ldots, and \quad cm_b = LB_i$$

$$then \quad cm_c = LC \tag{7}$$

For the navigation control:

$$R^{j}_{cn} \quad if \quad cn_a = LD_j, \ldots, and \quad cn_b = LE_j$$

$$then \quad y_j = f_j(y,u) \tag{8}$$

where $cm_a, \ldots cm_b$, cm_c, cn_a and cn_b are linguistic variables representing the process state variables and the control variable; $LA_i, \ldots, LB_i$, LC_i, LD_j and LE_j are the linguistic values of the linguistic variables $cm_a, \ldots cm_b$, cm_c, cn_a and cn_b in the universes of discourses $U, \ldots, V$, W, Q *and* G respectively, $i=1,2,\ldots,n$.

A conjunction of input variables associated with their respective linguistic values determines a linguistic value associated with the output variable. All rules are evaluated in parallel, and their outputs are combined to a fuzzy set, which has to be defuzzified to receive the crisp output value. The conjunction of the inputs is achieved by the *min*-operation, and for aggregating the outputs of the rules, the *max*-operation is utilised. Function *y* is used for the estimation of the parameters of the fuzzy-neuro controller system.

There exist three kinds of error signals in the proposed system. One of them is the difference between the desired response and the process output. This error e_l, called the learning error, is used to learn correct control action *u* and will tend to zero in the time interval of interest with increasing iteration number. The other error e_m, called the mission error, is the error between the mission requirements and the process statement. The last error e_{me}, called the measured control error and defined as the difference between the set-point and the process output, is primarily employed to construct the rule-base which will subsequently be used in the fuzzy control system.

4. CONCLUSION

This paper has described how fuzzy and fuzzy neural techniques are used for the development of the fuzzy model and fuzzy neural controller for the motion control of the GARBI underwater robot. The problems and difficulties with applying a model-following approach have been highlighted and an alternative intelligent systems approach has been presented.

The next phase of the work is to undertake an extensive simulation study in order to evaluate the motion control system prior to undertaking trials at sea.

Acknowledgement

The authors wish to acknowledge the support for this work provided by the British Council under the British/Spanish Acciones Integradas programme

REFERENCES

1. Amat, J.,Batlle, J.,Casals, A., & Forest, J. 1996. GARBI: The Low Cost ROV, Constraints and Solutions.
2. --- 1995. *GARBI: The UPC Low Underwater Vehicle.* Undersea Robotics & Intelligent Control, U.S/Portugal Workshop. pp. 19-24.
3. Amat, J.,Codina, J.,Cufi, X., & Puigmal, J. 1995. *Vision Based Control of Underwater Vehicle for Long Duration Observations.* ICAR' 95. pp. 273-277,
4. Takagi, T., & Sugeno, M. 1985. Fuzzy Identification of Systems and its Applications to Modeling and Control. *IEEE Trans. on Systems, Man and Cybernetics,* **SMC-15**, pp. 116-132.
5. Palman, R.,Driankov, D., & Hellendoom, H. 1997. *Model Based Fuzzy Control.* Germany: Springer-Verlag Berlin Heidelberg.
6. Mamdani, E. H. 1974. Application of Fuzzy Algorithms for Control of Simple Dynamic Plant. *Proc. of the IEE,* **121**, pp. 1585-1588.
7. Zadeh, L. A. 1968. Fuzzy Algorithms. *Information and Control,* **12**, pp. 94-102.
8. Terano, T.,Asai, K., & Sugeno, M. 1992. *Fuzzy Systems Theory & its Applications.* Academic Press, Inc.
9. Lee, C. C. 1990a. Fuzzy Logic in Control systems: Fuzzy Logic Controller Part I. *IEEE Trans on Systems, Man and Cybernetics,* **20**, pp. 404-418.
10. Takagi, H. 1990. Fusion Technology of Fuzzy Theory and Neural Networks - Survey and Future directions. *In P. of the Intr. Conf. on Fuzzy Logic and Neural Networks,* pp. 13-26.
11. Nauck, D., & Kruse, R. 1992. A Neural Fuzzy Controller Learning by Fuzzy Error Propagation. *Proc. NAFIPS'92,* pp. 388-397.
12. Sutton, R.,Taylor, H., & Roberts, G. N. 1997. Tuning Fuzzy Ship Autopilots using Artificial Neural Networks. *InstMC Transactions, Measurement and Control,* **19** (2), pp. 94-106.
13. Jang, J. & Sun, T. 1995. Neuro-Fuzzy Modelling & Control. *Proc. IEEE,* **83**, pp. 378-405.
14. Jang, J-S. R. 1993. ANFIS: Adaptive – Network - Based Fuzzy Inference System. *IEEE Trans on Systems, Man and Cybernetics,* **23** (3), pp. 665-685.
15. Nie, J. & Linkens, D. 1995. *Fuzzy-Neural Control Principles, Algorithms and Applications.* Prentice Hall.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Simplified Fuzzy Logic Control Scheme

C. M. Lim[a] and Y. Z. Ouyang[b]

[a]Department of Electronic and Computer Engineering, Ngee Ann Polytechnic
535 Clementi Road, Singapore 599489

[b]Hwa Chong Junior College
661 Bukit Timah Road, Singapore 269734

Abstract: A method to further simplify an existing fuzzy logic control scheme, which has nine linguistic control rules, is proposed. Within the framework of the proposed method, the resultant control scheme makes use of only three control rules. Consequently, the computation time required to find the proposed controller output is reduced. Most importantly, extensive simulation and real-time studies confirm that the proposed fuzzy logic control scheme is effective.

1. INTRODUCTION

The application of fuzzy logic control to enhance the performance of commercial products and industrial processes has received much attention in 1980s and, most importantly, encouraging results have been reported [1].

Motivated by the various successful applications of fuzzy logic control, one research area, which has attracted the attention of researchers, is in simplifying the structure of fuzzy logic controllers, FLC, without degrading the overall control performance. For instance, in [2,3], two or even three input signals to a FLC are linearly combined into one signal in order to reduce the number of control rules. While in [4,5], the non-fuzzy singleton type of membership function was proposed to reduce the computational time required to generate the final control signal.

Recently, a FLC with a low-dimensional number of fuzzy labellings, i.e., with a 3x3 partition, and using the "product-sum-gravity" method of reasoning [6], was proposed by Otsubo et al [7]. The FLC in [7] has been demonstrated to be effective, however, the only simulation results were presented. As such, there are rooms for improvement.

In the paper, the FLC reported in [7], is reexamined and simplified such that the number of fuzzy labellings used is reduced from a 3x3 partition to a 3x1. As a result, the computation time required to generate the control signal of the proposed FLC is less than that of the FLC reported in [7]. Most importantly, it is demonstrated using both simulation and experimental studies that the proposed FLC is effective for different systems.

2. PROPOSED FUZZY LOGIC CONTROLLER

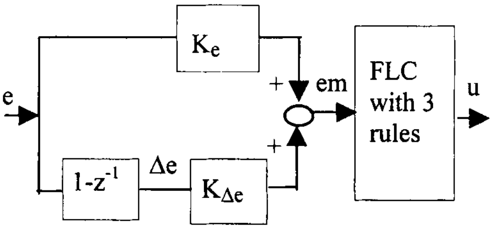


Figure. 1 Proposed FLC
(Note: u can be replaced by Δu to introduce an I-term).

Figure 1 shows the block diagram of the proposed FLC where e, Δe and u are the error, the change in error and the control signal, respectively. In order to reduce the number of control rules needed [3 , 4], the signals e and Δe are linearly combined to yield em which is given by

$$em = K_e e + K_{\Delta e}\Delta e \qquad (1)$$

where $(K_e, K_{\Delta e})$ are predefined constants.

Both signals em and u when fuzzified are defined to have 3 labels, viz., positive (P), zero (Z) and negative (N). As such the maximum of control rules required is only 3 and there are given by

R1: If em is P then u is P
R2: If em is Z then u is Z_h
R3: If em is N then u is N

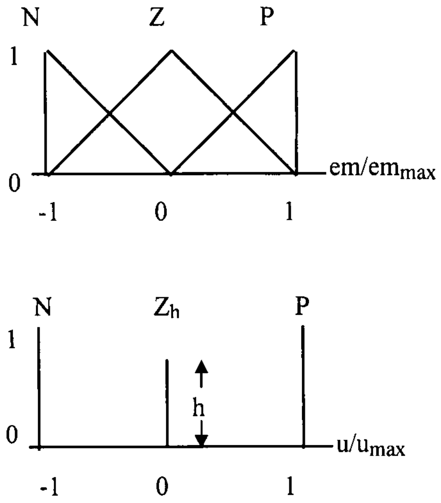


Figure 2: Membership functions

The membership functions of the labels are as shown in Fig. 2. Specifically, triangle-type of membership functions are chosen for the labels of em. However, singleton-type of membership functions are chosen for the labels of u. Of particular interest in Fig. 2 is the membership function for Z_h where $h > 0$. Such a membership function has been shown to yield results which are closed to human thinking [6]. Furthermore, by setting $h > 1$, Rule 2 is emphasised where as for $1 > h > 0$ the Rule is constrained [6]. When h is set to 1, the standard singleton-type of membership is obtained.

The output of the simplified FLC is obtained using the "product-sum-gravity" method of inferencing [6,7].

Remark 1: When an integral term is needed, the fuzzy controller output in Fig. 1 is to be replaced by $\Delta u(k)$ and the control signal u(k) to be applied to the plant is given by

$$u(k) = u(k-1) + \Delta u(k) \qquad (2)$$

Alternatively, an extra term I(k) can be added to u(k) of Fig. 1 to obtain the final signal, $u_f(k)$, as shown in Fig 3. Thus,

$$u_f(k) = u(k) + I(k) \qquad (3.1)$$

where

$$I(k) = I(k-1) + K_i\, e(k) \qquad (3.2)$$

and K_i is a constant.

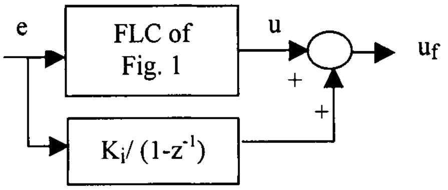


Figure 3: Proposed FLC with an I-term added.

Remark 2: In [7], e and Δe are used separately as inputs. Consequently, nine rules are required and, upon comparison with the proposed method described above, a longer computation time is needed to infer the control signal.

3. SIMULATION RESULTS

The effectiveness of the proposed FLC has been evaluated through simulation studies. The plant chosen for study is taken from [7] and has a transfer function, G(s), given by

$$G(s) = \exp(-s)/(1 + s) \qquad (4)$$

Figure 4 shows the closed-loop system responses for different values of h. From the results obtained, it can be seen that by a suitable choice of h the closed-loop system performance can be optimised.

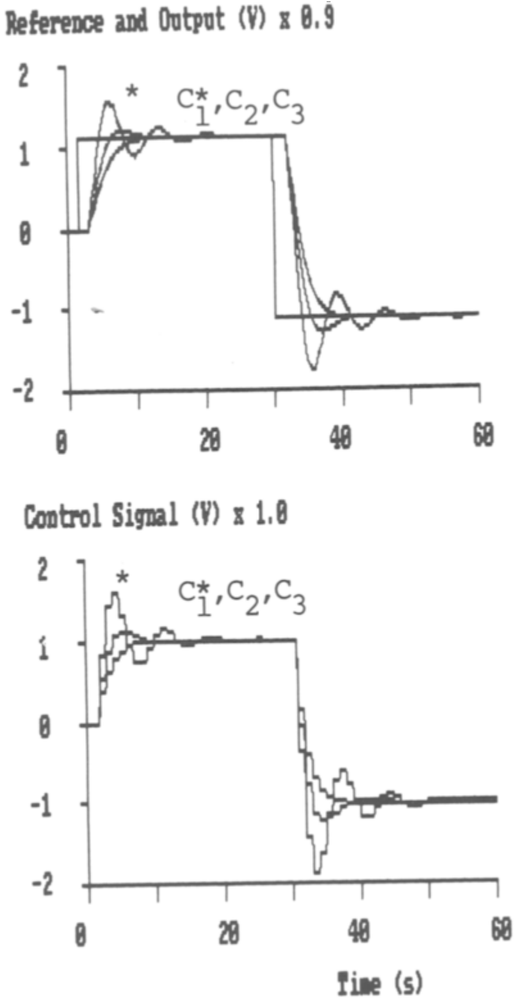


Figure 4. Simulation results (C1: h=0.5; C2: h=1.0; C3: h=1.5)

4. EXPERIMENTAL RESULTS

In order to demonstrate further the effectiveness of the proposed FLC, the setup as shown in Fig. 4 has been chosen for real-time control. The transfer function of the plant is given by

$$G(s) = K\omega_n^2/(s^2 + 2\varsigma\omega_n s + \omega_n^2) \qquad (5)$$

where (K, ς, ω_n) are its parameters whose actual numerical values are not needed in the design of the FLC. The output of the FLC is constrained within physical limits.

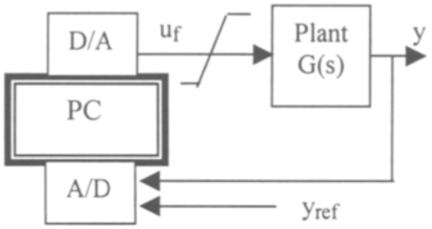


Figure 5. Experimental setup

The numerical values chosen throughout the experimental studies are as shown in Table 1 where T_s is the sampling period.

Table 1
Numerical values of FLC

$em_{max} = 1.0$ V	$u_{max} = 2.5$ V
$K_e = 1.0$	$K_{\Delta e} = 7.0$
$K_i = 0.1$	$\lvert u_f \rvert \leq 2.5$ V
$T_s = 10$ ms	$K = 1.0$

Figures 6 and 7 show typical samples of experimental results obtained with the FLC parameters given in Table 1. Figure 8 shows another set of results obtained with K and T_s set to 5 and 5 ms respectively.

The experimental results confirm that the proposed FLC is effective provided the value of h is properly chosen.

Remark 3. The observation made in this paper, viz, the performance of the closed-loop system can be optimised by a proper choice of h, is similar to that reported in [6,7]. However, in this paper, the number of rules required is only three while in [6] and [7] twenty five and nine rules are required respectively, and no real-time results were presented.

5. CONCLUSION

A simplified fuzzy logic control scheme has been presented and demonstrated to be effective using both simulation and real-time studies. Specifically, the proposed control scheme is just as effective as existing control schemes, however, the number of control rules required by the proposed scheme is minimal.

Future work to assist a user to tune the value of h automatically to yield optimal closed-loop performance is needed. Preliminary work, in which the position of a dc motor is controlled using the proposed scheme, indicates that the value of h can be self-tuned with the aid of a performance index. Further details will be reported in a separate paper.

REFERENCES

[1] T. Terano, K. Asai and M Sugeno, Applied Fuzzy Systems, AP Professional ,1994.

[2] C.M.Lim, Application of a new fuzzy logic control scheme for stability enhancement-simulation and experimental results, Proc. Int. Joint Conf . 4th IEEE Int. Conf. Fyzzt Syst. And 2nd Int. Fuzzy Engng Symp, Vol. 3, (1995) 1339-1348.

[3] C. M. Lim, Implementation and Experimental studies of fuzzy PID Controllers, Mechatronics, Vol. 4, No. 4, (1994).

[4] H. Ichihashi and H. Tanaka, PID fuzzy hybrid controller, Proc 4th Fuzzy System Symp. (1988) 97-102.

[5] M. Maeda and S. Murakami, Fuzzy control and its application, Systems, Control Inform., 34 , (1990), 282-287.

[6] M. Mizumoto, Fuzzy controls by fuzzy singleton-type reasoning method, Fifth IFSA World Congress (1993), 945-948.

[7] Otsubo et al, Improvement of control performance for low-dimensional number of fuzzy labelings using simplified inference method, Fuzzy Sets & Syst. 90, (1997), 37-44.

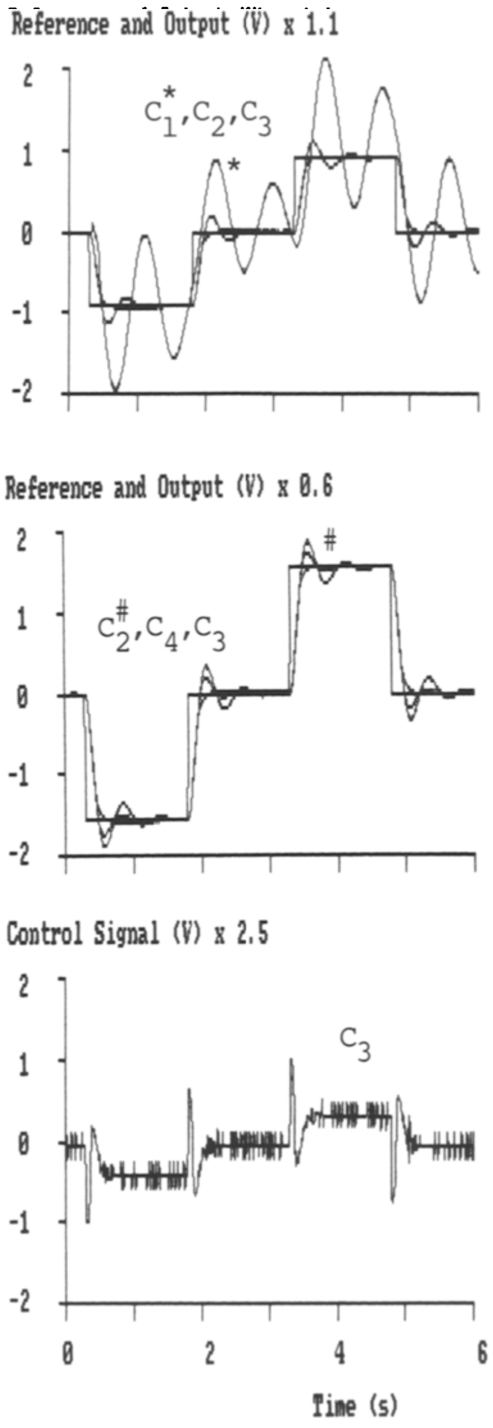


Figure 6. Experimental results (C1: without FLC; C2: h=1.7; C3: h=0.3; C4: h=1.0)

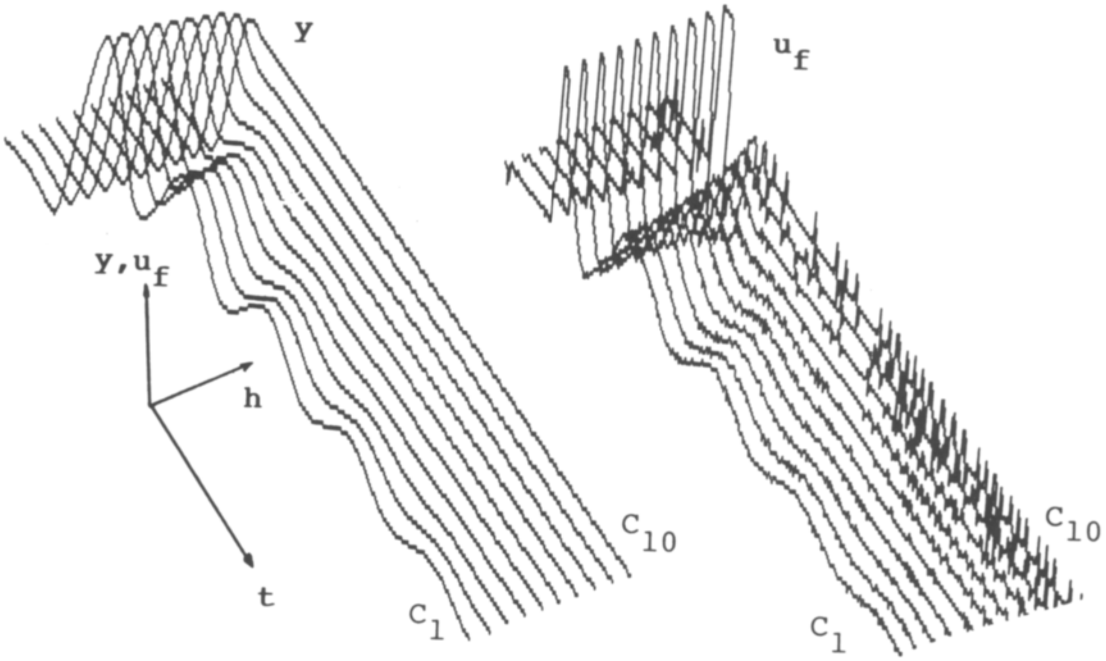


Figure 7. Experimental results – plant with nominal gain. (C_1: h = 2 and C_{10} : h = 0.2)

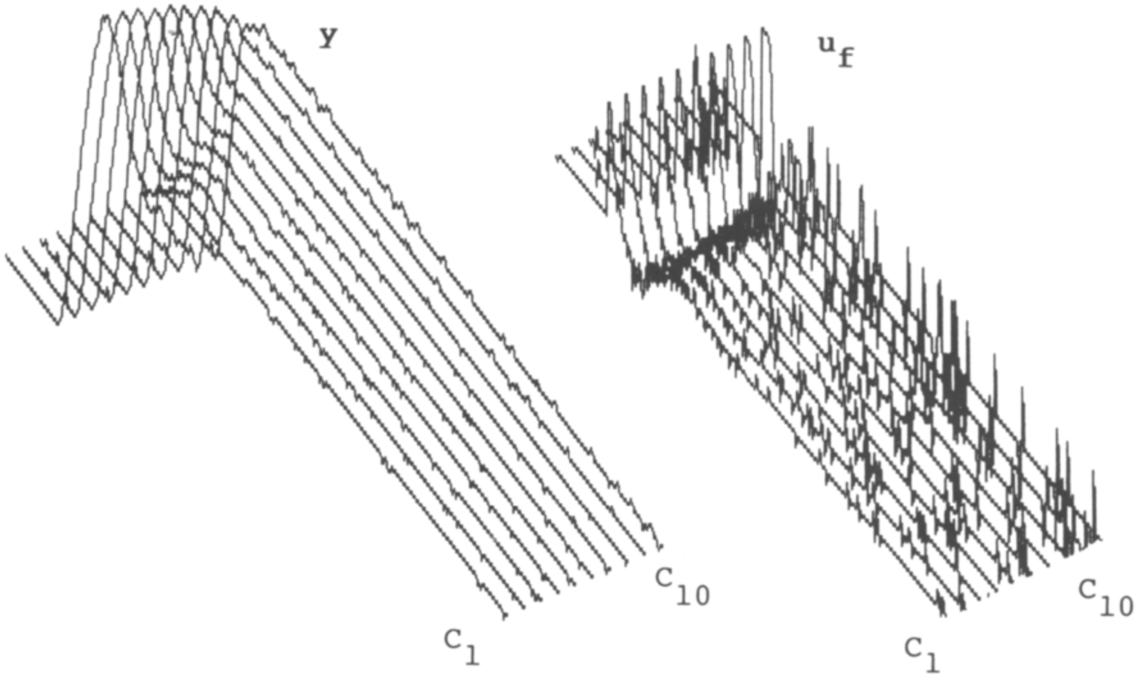


Figure 8. Experimental results – plant with high gain . (C_1: h = 2 and C_{10} : h = 0.2)

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

CONTROL STRATEGIES OF INJECTION DURING OPERATION OF S.I. ENGINES

Jorge J. G. Martins

Dpt. Engenharia Mecanica - Universidade do Minho
4810 Guimaraes - Portugal

The Spark Ignition Engines for road vehicles have very sophisticated control strategies, so the exhaust fumes can be efficiently treated in the so called "3-way" catalyst. To achieve an efficiency over 95%, the catalyst should be hot (over 400°C) and should receive an exact (1%) stoichiometric mixture.

In order to get a hot catalyst as early as possible (after a cold start) a heat transfer model was developed and the validation of the model was made with two different running engines.

To calculate the mixture strength during transients (accelerations) in cold or hot engines, another model was developed. This model consists of various sub-models, such as gas dynamics, transient fuel response (fuel deposition and evaporation from walls), and control logic's of the ECU of the engine. Comparisons between different fuelling logic's and sensor responses were made, after validation of the model with a running engine in an inertial dynamometer.

1. INTRODUCTION

Road vehicle related exhaust emissions legislation was first introduced in the USA and Europe in 1970. Since then the level of the permitted emissions have been reduced to almost negligible levels, comparing to pre-1970 motorcars. This resulted that, in parallel with safety, emissions are one of the most important areas in the vehicle development. One of the most important items in the emission control is the catalytic converter. In it the "dirty" exhaust gases clean themselves with a very high efficiency. The more common catalyst consists of a ceramic honeycomb with square cross sections of about 1mm covered with active catalytic material (noble metals such as Pt, Pd, Ru and Rh). With this design the exhaust gases spread over a large surface area where the chemical reactions take place.

Current and future tail pipe motorcar emissions legislation have compelled manufacturers to control engine parameters very effectively. The gases that are limited by legislation are the CO, HC and NO_x[1]. In the catalytic converter these gases react to form less harmful O_2, N_2, CO_2 and H_2O, which are then discharged into the atmosphere by the vehicle tail pipe. Nowadays engines need to burn an exact stoichiometric mixture, in order to allow the three-way catalyst to perform with full performance. Also, the catalyst needs to be above the "light-off" temperature to perform effectively.

From the above, we can conclude that, in order to achieve minimum exhaust emissions, one should assess the potential of reducing emissions by improving the fuelling strategies and reducing the catalyst warm-up period.

1.1. Fuelling Strategy

When the vehicle runs at steady-state conditions it is not too difficult to achieve a stoichiometric mixture. If something goes wrong, the oxygen.sensor and the closed loop control reinstate the stoichiometry[2]. The problem is to maintain the mixture strength during throttle excursions, specially when the throttle is opened quickly.

During a sudden throttle opening there are various factors that play an important role in the fuelling strategy, such as sensor response time, fuel deposition and evaporation, air flow and fuelling logic (enrichment). Even the conversions from analog to digital and vice versa in the control unit are important in the final result. All these factors have to be evaluated and optimized so the stoichiometry can be achieved at all times. Other important results can be obtained from evaluations of the fuelling strategy. How does a specific fuelling logic reacts to problems such as air leaks into the exhaust manifold, exhaust back pressure, etc.? These factors, although not common in normal conditions, can produce unexpected results in the engine.

1.2. Warm-up of Exhaust System and Catalyst

The catalytic converter induces the required oxidation and reduction reactions only if its temperature is above a certain level, usually 300°C-400°C. Only then the high conversion efficiency of the catalyst cleans the exhaust gases (from CO, HC and NO_x into O_2, N_2, H_2O and CO_2) up to 98%.

The normal way a catalyst is heated is by the passage of the hot exhaust gases. If its temperature allows it to induce chemical reactions (exothermal), they also contribute to the heating process. If the catalyst is placed near the exhaust ports of the engine and well insulated, it should do the job. However, if the catalyst temperature gets too high (above 1100°C) it will degrade, which shows the need for some sort of cooling.

It is not only the insulation of the exhaust pipes and of the catalyst itself that enhances the warm-up times. The lower the mass of the pipes, the lower the time for reaching the required temperature, by reduction of the thermal inertia of the system. Usually, insulation (commonly sprayed ceramic) increases the mass of the exhaust system, unless it uses very light materials or just an air pocket between concentric tubes.

There are other ways of heating-up the catalysts using external sources. The catalyst substract can be made of metal (with lower thermal inertia than the more current ceramic) and be electrically heated. Also, heat can be supplied by burning fuel upstream of the catalyst. During the first instants after a cool start, the engine needs a rich mixture to run properly. Injecting some air upstream of the catalyst forces oxidation reactions which produce heat to the catalyst.

1.3. Proposed Work

The areas of study were already selected as the fast warm-up of catalyst and the fuelling logic (particularly during throttle sudden openings). Various numerical models were written and various rigs were built to study and evaluate the selected areas of interest.

For the thermal study, two engines were connected to the dynamometer test bed and their exhaust systems were extensively monitored with thermocouples. A numerical model was developed to predict the behavior of various types of exhaust configurations. The laboratory tests were used to validate the model.

For the engine response study a special test bed was manufactured with the possibility of including inertia (to model the engine powering a vehicle). Various numerical models were written to calculate the response of the engine, such as air-flow, ECU logic, sensor response, fuel evaporation, etc. To study the fuel behavior (deposition and evaporation) other test rigs were built, where various fuelling conditions could be examined. The outcome of this study was implemented into the fuel evaporation model. A parallel work was also undertaken, to measure the temperature of the inlet valve. This would indicate the potential from the valve to promote fuel vaporization.

2. EXPERIMENTAL RIGS

This is a brief description of the various rigs used for the project.

2.1. Exhaust System Temperature Measurement

The dynamometer test bed was a conventional one and the temperature of the exhaust systems of the two engines was extensively monitored by thermocouples. The temperature was measured both at the gas side and at the wall side at different locations from near the exhaust valve to the catalyst.

The tests were performed with the engine starting from cold. Sometimes it was difficult to start the engine, as no enrichment was supplied. Various tests had to be repeated for this reason. Repetition of tests was a time consuming job, because the engine had to cold down completely before re-starting again.

2.2. Acceleration Tests

An inertial dynamometer test bed was used to model the acceleration behavior of a vehicle in second and forth gear, during a full throttle, fast acceleration. The inertia was implemented in the test bed my adding flywheels to the engine/dynamometer shaft[3]. A small flywheel modeled 2nd gear acceleration, while a large wheel was used as 4th gear inertia. The acceleration patterns were from light load to full throttle in 0.3 second and a longer one, with 3 second duration. Tests were carried on with the engine coolant set at 90°C and at 30°C. The resulting mixture strength was measured by a fast response AFR meter at the exhaust gases.

With this test rig it was possible to optimize the fuel enrichment levels both for Single Point Injection Engines and for Multi Point Injection Engines. These tests were also used to validate the developed model.

3. MODELS

Various numerical models were developed to perceive the behavior of the engine as far as exhaust emissions are concerned. Furthermore, the catalyst should be hot to induce the reduction of the harmful gases from the exhaust, so the better way to heat-up the catalyst as soon as possible was assessed.

3.1. Exhaust System Thermal Modelation

The exhaust system was modeled as a succession of lumped sections. For each section the various processes of heat transfer were applied for the tube and gas. The following equations were used to produce the model.

3.1.1. Heat Balance on Exhaust Gas (Figure 1)

$$m_g Cp_g \frac{dT_g}{dt} + \dot{m}_g \left(H_{g,j} - H_{g,j-1}\right) = U_i A_{ti} \left(T_t - \overline{T}_g\right) \quad (1)$$

from which

$$\frac{dT_g}{dt} = \frac{U_i A_{ti}}{m_g Cp_g}\left(T_t - \overline{T}_g\right) - \frac{\dot{m}_g}{m_g}\left(T_{g,j} - \overline{T}_{g,j-1}\right) \quad (2)$$

where

$$U_i = \frac{1}{\dfrac{1}{Eh_g} + \dfrac{d_{ti} \ln\left(\bar{d}_t / d_{ti}\right)}{2k_t}}$$

and $\bar{d}_t = \sqrt{\left(\dfrac{d_{to}^2 + d_{ti}^2}{2}\right)}$

$$h_g = \frac{k_g}{d_{ti}} 0.024 \,\mathrm{Re}_g^{0.8} \,\mathrm{Pr}_g^{0.5}$$

$$E = 2$$

3.1.2. Heat Transfer on Tube and Shell Wall (concentric tube arrangement) (Figure 2)

Assuming that radial temperature differences in the tube and shell walls are small compared with the temperature differences between exhaust gas and air, a heat balance on the tube gives:

$$m_t Cp_t \frac{dT_t}{dt} = A_{ti} U_i \left(\overline{T}_g - T_t\right) + A_{to} F_{t-s} \sigma \varepsilon_t \left(\widetilde{T}_s^4 - \widetilde{T}_t^4\right) + \frac{2A_{to} k_g \left(T_s - T_t\right)}{d_{si} - d_{to}} \quad (3)$$

for shell

$$m_s Cp_s \frac{dT_s}{dt} = A_{so} h_a \left(\overline{T}_a - T_s\right) + A_{si} F_{s-t} \sigma \varepsilon_s \left(\widetilde{T}_t^4 - \widetilde{T}_s^4\right) + \frac{2A_{to} k_g \left(T_t - T_s\right)}{d_{si} - d_{to}} + A_{so} \sigma \varepsilon_s \left(\widetilde{T}_a^4 - \widetilde{T}_s^4\right) \quad (4)$$

where

$F_{t\text{-}s}=1$, $F_{s-t} = \dfrac{A_{to}}{A_{si}} = \dfrac{d_{to}^2}{d_{si}^2}$

h_a – air side heat transfer coef.

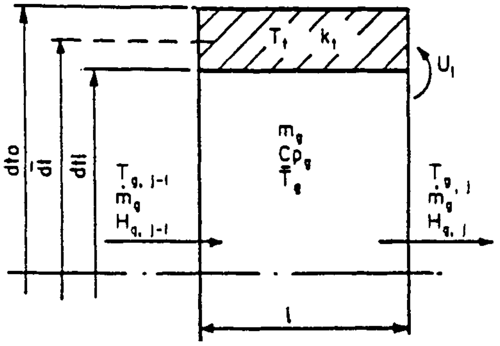


Figure 1 – Gas side

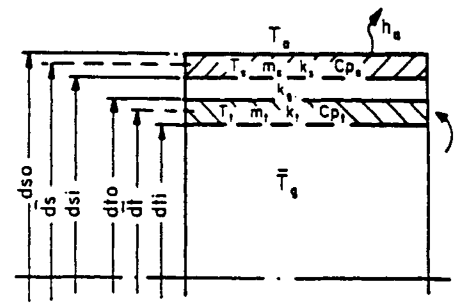


Figure 2 – Tube and shell

3.1.3. Other tube geometries

Two other tube geometries were modeled: insulated tube and concentric tube with insulation in the gap. The heat transfer equations are of the same type as (3) and (4).

With the described model it was possible to evaluate the potential for decreasing the time for catalyst heating. As an example, Figure 3 shows the result of the exhaust pipe (tube and shell wall type) during the first 100 seconds of the 75 US Federal Driving Cycle. The section shown is the last before the catalyst. The exhaust gas enters the catalyst at almost 500°C, at the end of that period. As it can be seen, the outer wall is always at low temperature, as a result of the resistance of the air gap between inner and outer walls. The inner pipe is made of thin stainless steel.

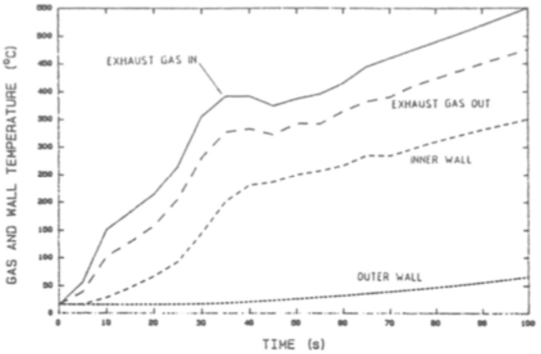


Figure 3 – Temp. vs. time at the entrance of catalyst

At the moment the work is continuing on the program for the catalyst, which should model the various chemical reactions and the various forms of fast warm-up.

3.2. Engine Response Modeling

3.2.1. Gas Dynamics Model

The NEL engine simulation model[4] was used to generate predictions of the unsteady gas flow in the engine intake and exhaust systems so the transducers see a true simulation of the engine flows.

3.2.2. Transfer Functions

In a running engine the values of mass flow, throttle position and manifold pressure are measured by instruments which feed the ECU of the engine with voltage inputs. To model the response of the engine, these signals need to be calculated with transfer functions representing the sensors. The output signals are, therefore, filtered versions of the input signals, connected to the calibration curves of each sensor.

The transfer functions consist of differential equations converting input (X_i) into output values (X_o) using a time constant T:

$$T\frac{dX_o}{dt} + X_o = X_i$$

More complex functions can be used, which include other aspects such as hysteresis.

3.2.3. Fuel Flow Model

In a multipoint injection (MPI) engine the fuel is prayed near (or over) the inlet valves. Part of the fuel is deposited on the walls, part is deposited onto the hot inlet valves (vaporizing fast [5]) and the remainder vaporizes as the injection takes place. The part deposited on the walls (D) forms a liquid pool which evaporates and entrains at a rate (τ) proportional to the stored liquid. On leaving the pool, the fuel is assumed to mix instantly with the air in the port (Figure 4). The basic model was based on the work of Aquino[6] and incorporated an air flow model associated with the fuel flow model.

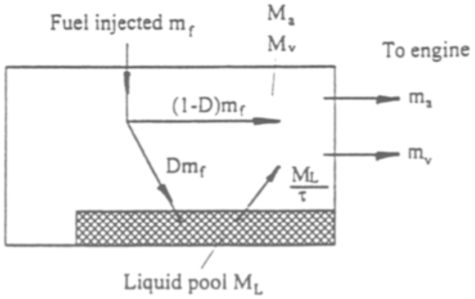


Figure 4 – Engine Evaporation Model

Mass continuity (fuel) within the manifold gives:

$$\frac{dm_L}{dt} = D\dot{m}_F - \frac{m_L}{\tau} \quad \text{(for the liquid - L)}$$

and

$$\frac{dm_V}{dt} = (1 - D)\dot{m}_F + \frac{m_L}{\tau} - \dot{m}_{Ae}\left(\frac{m_V}{m_A}\right)$$

(for the vapour - V)

However, when the inlet valve is open, the shearing stress of the high velocity air passing over the fuel patch carries fuel with it (the deposition is near the inlet valve). Therefore the model was modified to incorporate the shearing stress, to limit the quantity of fuel stored in the manifold, creating a wall film flow, calculated as a Couette flow and an interfacial shear stress from the Blasius friction factor. The results of the model were compared with those from the test rig. Two types of measurement were made: torque and air-fuel ratio (AFR). The torque was calculated by measuring flywheel velocity and acceleration and the AFR was measured using a fast response meter.

The results can be seen in Figure 5, where measured and predicted values of engine torque response during a fast throttle opening (0.3 s) can be seen, for an initial engine speed of 1570rpm, fourth gear inertia and warm engine. The best fit was for D=0.6 and $1/\tau=3.3$ for this test and for others. When the engine temperature falls to 30°C, the $1/\tau$ falls to 0.4^3.

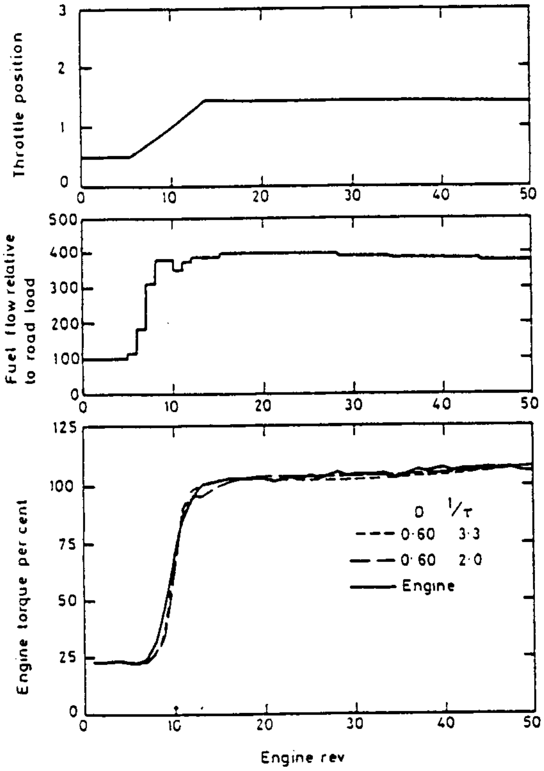


Figure 5 – Comparison of measured and predicted torque

3.2.4. Engine Response Model

With the fuel model validated, modifications were introduced to link it to the air-flow model based on the air dynamics calculations. The full model incorporates (Figure 6):

- gas dynamics, to calculate the air flow inside the inlet manifold and engine;
- sensor response, to calculates the data introduced into the engine ECU;
- ECU fuelling logic, leading to the fuel injections;
- fuel model, leading to the AFR entering the engine.

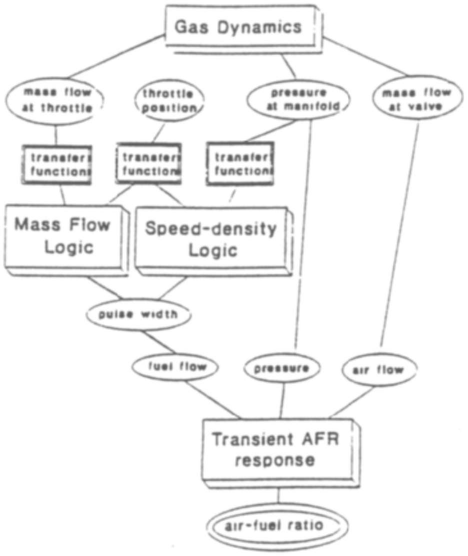

Gas Dynamics
mass flow at throttle
throttle position
pressure at manifold
mass flow at valve
transfer function
transfer function
transfer function
Mass Flow Logic
Speed-density Logic
pulse width
fuel flow
pressure
air flow
Transient AFR response
air-fuel ratio

Figure 6 – Flowchart of the numerical model

3.2.5. Modelization of the ECU

The electronic control unit (ECU) of the engine converts the results of the engine sensors into injection pulse widths (time for injection). There are two systems for that calculation:

- massflow - the air flow is measured (e.g. by a hot wire) upstream of the throttle;
- speed-density - the air flow through the inlet valves is calculated from the manifold pressure and engine speed.

During steady-state operation both systems read the same, but during throttle openings the first systems reads more mass that the speed-density. The reason lies on the location of the measurement: throttle plate for the massflow system and inlet valves for the other. In fact, during a throttle opening, more air passes through the throttle than through the valves, because the volume between both locations has to be filled with air from a low pressure (light load, closed throttle) to atmospheric pressure. Furthermore the fuel could be injected at different timing (during open or closed valve periods), the injection could be parallel, grouped or sequential and during the throttle opening or closing moments. It is known that injection on open valve is good for throttle response and injection on closed valve enhances mixture preparation. The model allows to calculate the engine response (AFR and torque) at different operating conditions (speed, temperature, throttle opening speed, measuring logic, etc.).

Results can be seen in Figure 7 for fast (0.3 s) throttle opening with warm (90°C) engine.

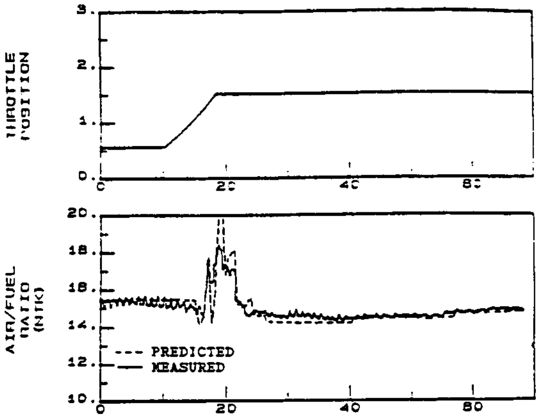


Figure 7 – AFR values for measurement and prediction

The optimum fuelling requirements was calculated with the model, so the AFR did not oscillate more than 3 (Figure 8) at any condition (the basic ECU logic gave excursions of over 8 AFR - Figure 9). Both figures are for 1500rpm, coolant at 30°C, and fast (0.3 s) transient, which were the more severe tests done.

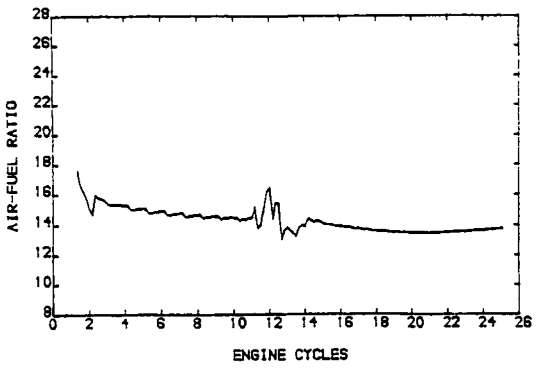


Figure 8 – Optimum AFR response

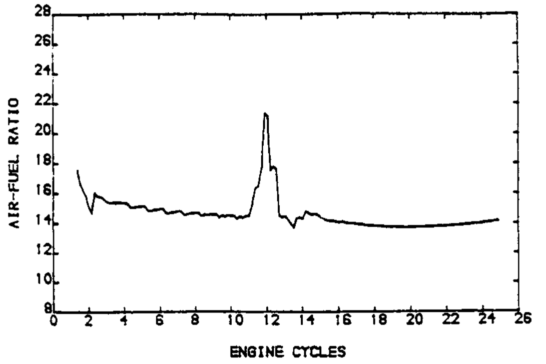


Figure 9 – Response of the engine ECU

ACKNOWLEDGEMENT

I would like to thank the National Engineering Laboratory in East Kilbride, U.K., where all tests were performed.

REFERENCES

[1] MARTINS, J.J.G.
Emissions from Internal Combustion Engines: European Legislation and Impact upon Technology, in 'Industrial Air Pollution – Assessment and Control' ed. A. Muezzinoglu and M.L. Williams, pp 175-184, Springer-Verlag, 1992

[2] HEYWOOD, J.B.
Internal Combustion Engine Fundamentals, McGraw-Hill, 1988

[3] BOAM, D.J., FINLAY, I.C. and MARTINS, J.J.G.
A Model for Predicting Engine Torque Response during Rapid Throttle Transients in Port-Injected Spark-Ignition Engines, SAE Transactions, Journal of Engines, pp 999-1010, Vol 98, 1989

[4] BINGHAM, J.F.
Intake System Design Using a Validated Internal Combustion Engine Computer Model, International Conference on Computers in Engine Technology, Cambridge, paper C25/87, Mechanical Eng. Publ. Ltd, 1987

[5] MARTINS, J.J.G. and FINLAY, I.C.
Fuel Preparation in Port-Injected Engines, SAE Transactions, Journal of Fuels & Lubricants, pp 621-632, Vol 101, 1992

[6] AQUINO, C.F.
Transient A/F Control of the 5-liter Central Fuel Injection Engine, SAE paper No.810494, 1981

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A FLEXIBLE END-EFFECTOR FOR ROBOTIZED ASSEMBLY

J.A.N. Loureiro[a] and M.P.S.F. Gomes[b]

[a] Dep. de Engenharia Mecânica, Instituto Politécnico da Guarda,
6300 Guarda, Portugal (To whom all correspondence should be sent).
jloureiro@ipg.pt

[b] Dep. de Produção e Sistemas, Universidade do Minho,
4800 Guimarães, Portugal
pgomes@ci.uminho.pt

Abstract: This paper presents requirements for robotized assembly, and a possible solution coined CTFMR (end effector for flexible robotized assembly). The CTFMR was designed, integrated with a PUMA 560 C robot, and tested by the first author, in the CIM laboratory of the IPG (Polytechnic Institute of Guarda), Portugal.

The CTFMR includes modules aimed at the safety of people and equipment, continuous error measuring between peg and hole for active insertion, and automatic tool change. A new pneumatic self-centering three-fingered gripper is proposed. This gripper is capable of detecting the absence/presence of objects between the fingers using optical fibre sensors, and of grasping objects of different shapes and sizes, either from the inside or outside of the object. It also prevents the objects from slipping by automatic increase of the gripping forces.

The main interest of the CTFMR concept is the increase of safety, flexibility and productivity of robotized assembly.

Keywords: Robotized assembly, active insertion, compensation mechanism, gripper, flexible end-effector for robotic assembly.

1. INTRODUCTION

The paper has the following objectives:

- to identify the causes of low application of robots in industrial assembly,
- to report on the design, integration and test of devices which can increase the flexibility and productivity of robots used in assembly.

Robots can advantageously replace people at routine, painful and dangerous tasks, assuring the quality, consistency and hygiene of products. The need to use robots in assembly is felt in industry, especially in the electrical, electronic and automotive sectors.

There are ethical, economic and technical reasons that restrain the expansion of robots in industrial assembly:

- potential unemployment caused by a decision to automate,
- economic aspects related to investment and revenues in robotized batch production when compared to manual work or to the use of assembly machines,
- technical limitations of robots related to using generic tools, to previous planning and programming, as well as to execution monitoring and recovering from failures.

In order to perform a task, a robot must be equipped with an end-effector (tool, gripper or dextrous hand), and programmed to locate and orientate its TCP (Tool Center Point) towards the target as well as to actuate the end-effector and other peri-robotic equipments (part feeders, fixturing, etc). If the robot is provided with external sensors and algorithms for intelligent operation it will be able to work autonomously despite environment changes.

The robot's ability to carry out different tasks depends on its specifications (DOF's, accuracy, repeatability, reach, mobility, position and force control) and also on the:

- adaptation of the end-effector to several tasks, or automatic tool change which is faster than manual substitution,
- intelligent reaction to environment changes which is necessary in unattended shifts,
- communication with other computer control equipments (DNC computer, other robots and CNC machine tools).

2. THE STATE-OF-THE-ART IN ROBOTIZED ASSEMBLY

Research areas relevant to the application of robots in industrial assembly comprise:

- insertion strategies: passive (RCC or Remote Centre Compliance [6]), and actives (iRCC or instrumented RCC, wrist force sensors, micromanipulators [6] and coordinate control of multiple manipulator systems [3]),
- development of feeders and end-effectors (grippers, tools and dextrous hands) [2], [10],
- peg-in-hole insertion model [7],
- integration of sensors, controllers and actuators,
- classic control systems (P, PD, PI, PID-proportional integral derivative) together with neural and fuzzy controls [10],
- monitoring and control for recovery from failure,
- automatic methods for task planning and programming of robots and dextrous hands [1], [4], [9], [10].

3. THE CTFMR CONCEPT

The CTFMR device represents a mechatronics approach to improve quality, productivity and flexibility of industrial robots. Although devices for robotized assembly, safety and tool change mechanisms already exist (see section 2), the concept CTFMR drawns on all these and it also comprises a highly versatile new gripper. Several hardware and software components are necessary to integrate the CTFMR with a robot system. A PUMA 560 C robot was used in this work.

3.1. Hardware

Devices permanently attached to the robot's wrist:

- electrical device for safety and collision avoidance (Z axis),
- iRCC (instrumented RCC) for continuous error measuring durig insertion (X and Y axes),
- manual locking device of the iRCC if necessary,
- electrical and pneumatic interface for automatic tool change of end-effectors.

Equipment on the assembly table:

- a vertical tool magazine equipped with a suction module and a gripper. The gripper is pneumatically actuated, parallel, with three fingers for self-centering grasping of objects of different shapes, sizes and orientations. Two fingers are manually adjustable for positioning while the third is pneumatically actuated by programmable forces. The gripper detects the absence/presence of objects between the fingers using optical-fibre sensors and it also prevents the parts from slipping by automatically increasing the grasping force, under software control and the action of a proportional pneumatic pressure regulating valve,
- pneumatic self-centering fixture system for gripping the hole component during peg insertion,
- feeder for cylindrical pegs.

Other accessories:

- 2 power supply units (24 V DC),
- one infrared (IR) light barrier for robot perimeter control,
- 3 IR proximity sensors and 6 optical fibres in the gripper,
- inductive proximity sensors (in the magazine and feeders),
- direction-control valves with actuation by solenoid and pilot valve for the cylinders which actuate the tool change, fixturing and the right finger of the gripper,
- a proportional pneumatic pressure-regulating valve for the control of the grasping forces of the right finger,
- a PLC (Programmable Logic Controller) for analogue data acquisition and A/D convertion (iRCC potentiometers) and D/A convertion for control of the proportional valve.

Figures 1 and 2 show devices permanently attached to the robot's wrist and a close up of the CTFMR.

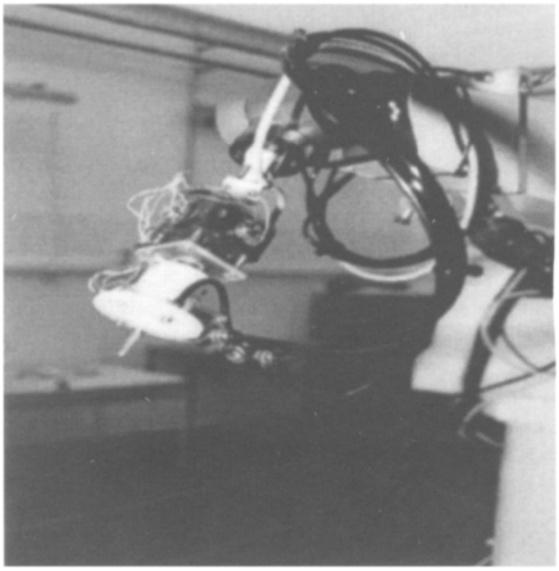


Figure 1: Devices attached to the robot's wrist.

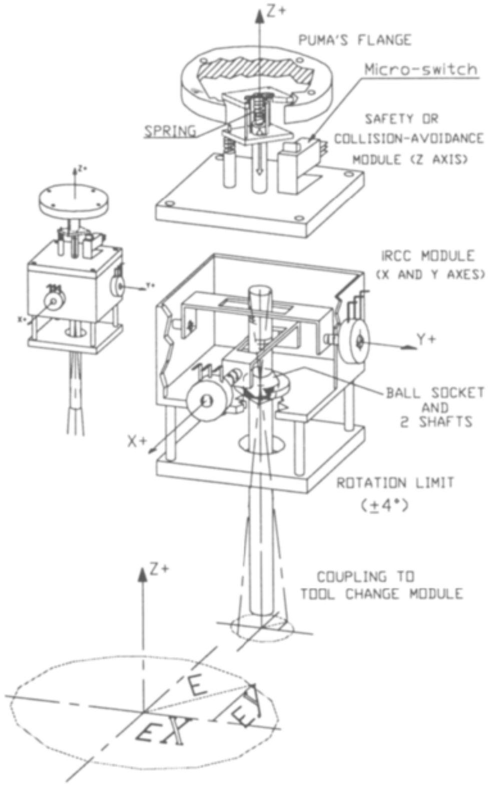

Z+
PUMA'S FLANGE
Micro-switch
SPRING
SAFETY OR COLLISION-AVOIDANCE MODULE (Z AXIS)
IRCC MODULE (X AND Y AXES)
Y+
BALL SOCKET AND 2 SHAFTS
X+
ROTATION LIMIT (±4°)
Z+
COUPLING TO TOOL CHANGE MODULE
E
EX
EY

Figure 2: A close up of the CTFMR (without tool change module).

Figure 3 shows the Tool Magazine. Fingers' movements for versatile and self-centering grasping of objects are shown in figure 4.

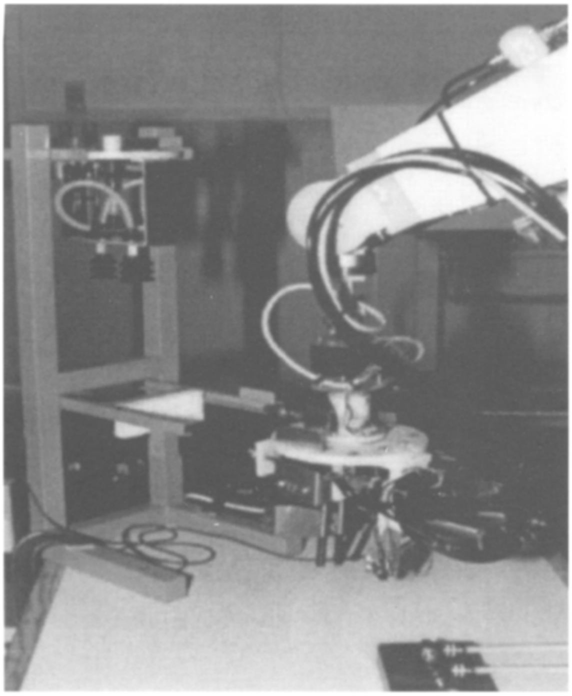


Figure 3: Tool Magazine with suction module (on top) and the new self-centering gripper.

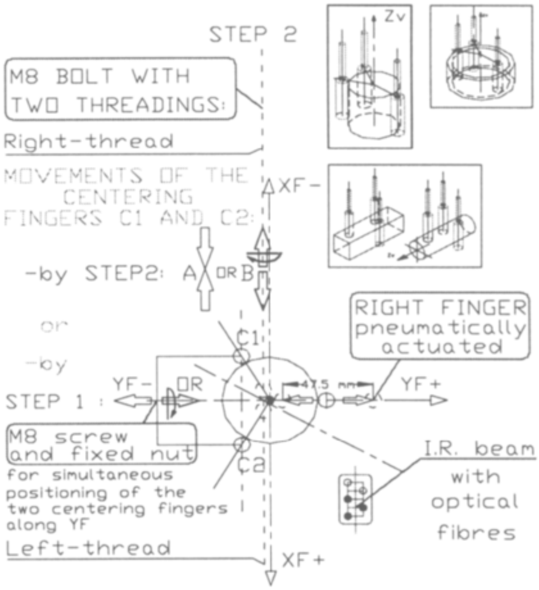

STEP 2
Zv
M8 BOLT WITH TWO THREADINGS:
Right-thread
MOVEMENTS OF THE CENTERING FINGERS C1 AND C2:
XF-
-by STEP2: A OR B
or
-by
STEP 1 :
YF-
OR
47.5 mm
YF+
RIGHT FINGER pneumatically actuated
C1
C2
M8 screw and fixed nut
for simultaneous positioning of the two centering fingers along YF
I.R. beam with optical fibres
Left-thread
XF+

Figure 4: Working principle of the proposed pneumatic self-centering three-fingered gripper.

3.2. Software

VAL II (Versatile Assembly Language)

Programs were developed in VAL II to control the end-effector, and the robot arm for active insertion based on data pre-processed by the PLC and supplied to the PUMA controller (master) for error correction during insertion.

STL (Statement List)

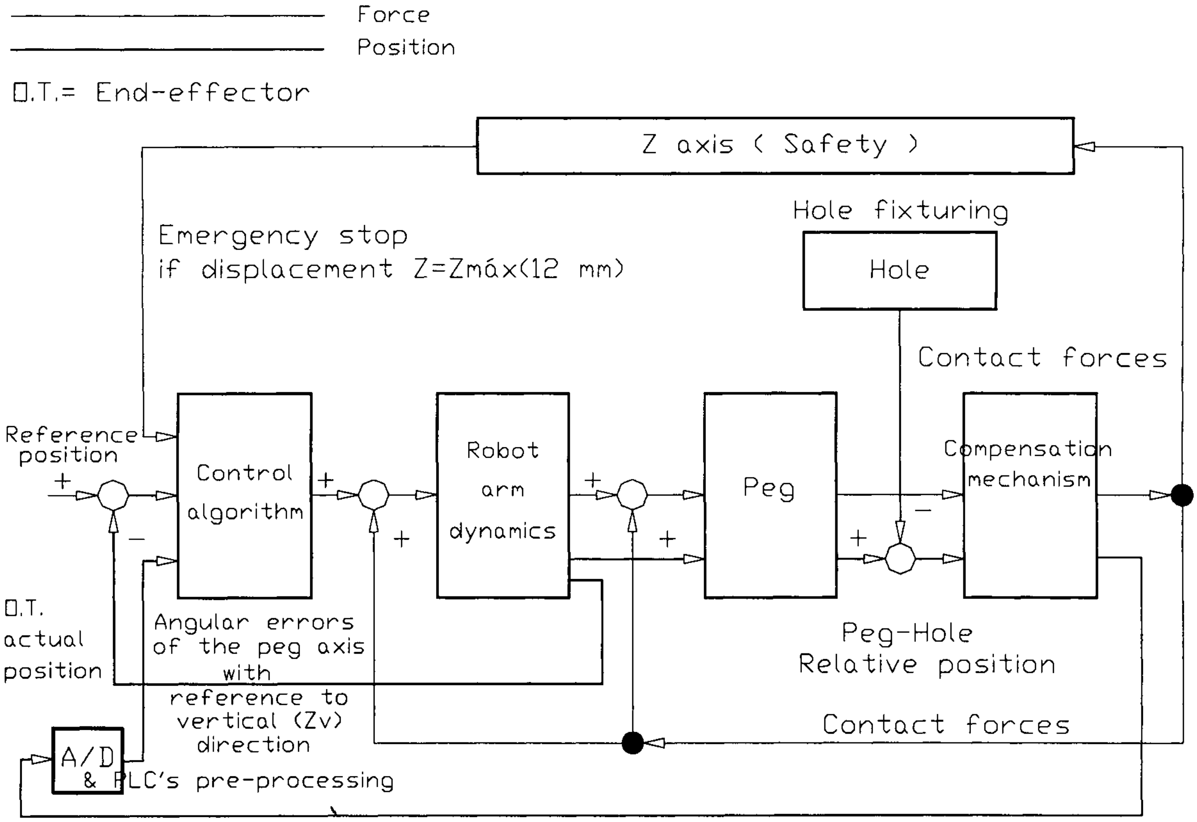
Programs were developed in STL to allow the PLC to perform:

- continuous measuring of the angle of the peg axis relative to the vertical direction by using two potentiometers and A/D convertion. This data is used by the PUMA controller for the active correction of orientation errors during insertion,
- detection of absence/presence of objects between the fingers of the gripper equipped with I.R. sensors and optical-fibres,
- automatic increase of the grasping force in case of slipping by means of greater pressure supplied to the right finger pneumatic actuator according to the output pressure of the proportional valve commanded by the PLC.

Figure 5 shows a diagram of the control loop.

4. CTFMR SPECIFICATION

The mechanical configuration, the working principle, as well as the integration of sensors, controllers and actuators for CTFMR operation are shown in detail and explained in **[5]**. The specifications of the CTFMR are summarised in Table 1.


Force
Position
D.T.= End-effector
Z axis (Safety)
Hole fixturing
Emergency stop
if displacement Z=Zmáx(12 mm)
Hole
Contact forces
Reference position
+
Control algorithm
+
Robot arm dynamics
+
Peg
Compensation mechanism
−
+
+
+
−
D.T. actual position
Angular errors of the peg axis with reference to vertical (Zv) direction
Peg-Hole Relative position
Contact forces
A/D & PLC's pre-processing

Continuous measuring of the angle of the peg axis relative to vertical direction-Zv (X and Y axes /potentiometers)

Figure 5: Block diagram of the assembly system using CTFMR.

Table 1:
Specifications of the CTFMR for use with industrial robots similar to the UNIMATION PUMA 560C

1 SAFETY BARRIER	4mx4m square around robot PUMA 2 sides: physical barrier 1m tall 1 side: I.R. reflex light barrier FESTO SOE-RT-Q-PS/O-K-LED 1 side: I.-R. Thru-beam OMRON (E3JK-5L/emitter and -5DM1/ receiver) signals: for 2 inputs of J147 of PUMA controller/ reaction by safety confirmation [8]
2 SAFETY OR COLLISION-AVOIDANCE MODULE (Z AXIS)	microswitch 24 VDC maximum length before actuation: 12mm mass: 0.058 kg signal: for input of J147 of PUMA controller/ reaction by safety confirmation [8]
3 COMPENSATION MODULE (X AND Y AXES)	crossed axes, centering springs, ball socket and 2 shafts 2 potentiometers, linear (50 kΩ) with rotary axes rotation limited to ±4° relative to vertical direction locking device: manually operated mass: 0.220kg pre-processing: A/D convertion and programming (PLC FESTO FPC 101-AF)
4 AUTOMATIC TOOL CHANGE MODULE	pneumatic interface: 2 plug/male connector snap-in type electrical interface: plug of 9 pins connector mass: 0.164 kg
5 TOOL MAGAZINE	type: fixed, vertical mounted with 2 horizontal U-shaped tracks per cell sensors: inductive proximity sensor OMRON TL-X1R5C1-GE signal: for input of J147 of PUMA controller
6 SUCTION MODULE	pneumatic interface: 2 socket/female connector snap-in type electrical interface: socket of 9 pins connector vacuum generation: by compressed air supplied to the ejector/Venturi (MECMAN 735-030 000-1) / exhaust: silencer vacuum ON/OFF: by external 3/2 way solenoid valve with normal position closed suction cups: 2 bellows suction cups (4.5 N) payload: 0.625 kg (vertical axis) mass: 0.569 kg
7 VALVES	directional control valves: 4/2-way, 5/2-way and 5/3-way valve with mid position closed/spring centered proportional pressure regulating valve FESTO MPPE 3-1/8-6 (0-6 bar)
8 PARALLEL SELF--CENTERING GRIPPER	fingers: 3 vertical cylinders with Ø9x30 mm (end cover) movements: 2 fingers with simultaneous motion for manual centering (Step1 and 2) 1 (right) finger pneumatically actuated (MECMAN 132 Ø10X50 mm) gripping forces: programmable from 8.6 N up to 47.1 N (with installed cylinder and 6 bar) pneumatic interface: 2 socket/female connector snap-in type electrical interface: socket of 9 pins connector range of manipulation: inner, outer (horizontal or vertical) always unilaterally outer grasping of cylindrical parts with vertical axis: 5≤Ø≤100 mm outer grasping of " " horizontal axis: 1.5≤Ø≤78 mm outer grasping of parts with two parallel faces: 1.5≤ • ≤78 mm inner grasping of parts with holes which diameter: 22≤Ø≤120 mm absence/presence and slipping detection by 3 I.-R. OMRON E3X-A41(NPN), 6 optical fibres E32-TC 200 and STL &VAL II programming payload: 0.8 kg (reccomended for PUMA 560 C) mass: 0.993 kg
9 PROGRAMMABLE LOGIC CONTROLLER	FESTO FPC 101-AF (slave controller) with: 21 Inputs/14 Outputs (digital interface) 8 I / 2 O (analogue interface)
10 PUMA 560 C ROBOT	Master contoller with standard J147 (16 I/16 O) without J149 or analogue KIT

5. RESULTS

The procedure used for the experimental evaluation of the CTFMR included two types of tests:

- a single test on purchased components and on each developed module working separately,
- general or integration test on the system first step by step and finally in automatic execution of robotized insertions.

Results presented in [5] relate to the vertical insertions of pegs performed by the PUMA 560C robot equipped with CTFMR using 0.5mm clearance between peg and hole (SKF 6206 bearing). The success in insertion was 93% and the average time of insertion and active corrections was 4.1 seconds.

Figure 6 shows insertion test and suction module operation.

6. CONCLUSIONS

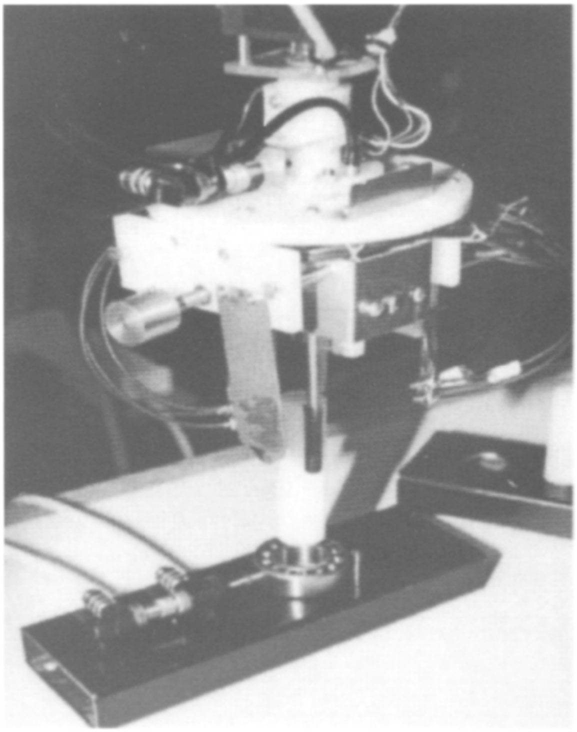
The CTFMR was developed to be integrated with any industrial robot with similar specifications to the UNIMATION PUMA 560C and to increase the safety, flexibility and productivity of robotized assembly.

Some limitations of this prototype were identified during tests, namely:

- misalignement between the interface for automatic tool change and each end-effector positioned in the magazine cell. This drawback can be overcome by using extra sensors and actuators in the magazine to improve coupling,
- low payload (0,8 kg) due to the mass of some gripper components. Possible solutions can be the selection of lighter materials and size reduction without compromising mechanical strength,
- some electrical components are prone to mechanical wear and tend to limit operation life. Optical transducers can replace potentiometers and the microswitch,
- the compensation mechanism only works correctly in vertical insertions with a maximum angular error of ±4° according to design solution.

The validity of the CTFMR concept was experimentally demonstrated with the prototype.

Priority was given to operation, with improvements being left for further developements.



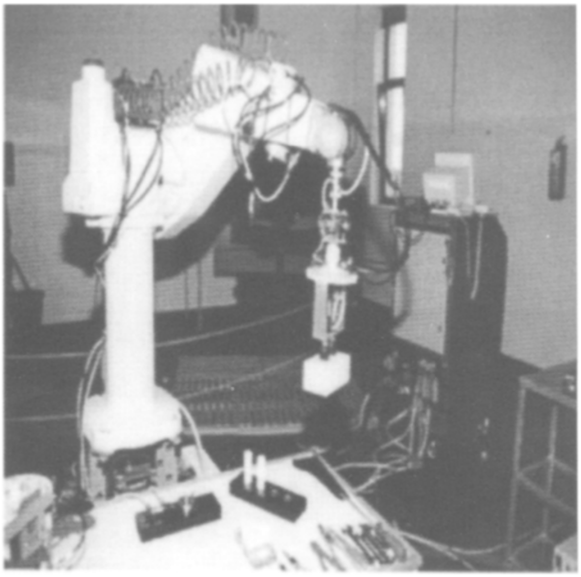


Figure 6: Tests performed with the new gripper and suction module.

The CTFMR allows the manipulation of objects of different shapes and sizes as shown in Table 1, the accomplishment of active insertions, and automatic tool change. Furthermore, it can have two potential uses:

- geometry acquisition useful for CAD/CAM (reverse engineering) or for quality control of parts using an inspection probe guided by the robot,
- difficult couplings thanks to the self-centering action based on the design of the proposed gripper.

ACKNOWLEDGEMENTS

The authors would like to thank the support of IPG, Prodep-II, and the students and technicians' collaboration in the experimental work.

REFERENCES

[1] BEKEY, G.A.; Liu H.; Tomovic R.; Karplus W.J.; *"Knowledge-Based Control of Grasping in Robot Hands using heuristics from Human Motor skills"*, IEEE Transactions on Robotics and Automation, pp 709-721, vol 9, n.6, December 1993

[2] HIRZINGER G.; Brunnen B.; Dietrich J.; Heindel J.;*"Sensor-based space robotics-ROTEX and its Telerobotic features"*, IEEE Transactions on Robotics and Automation, pp 649-663, vol 9, n.5, October 1993

[3] HSU P.; "*Coordinated Control of Multiple Manipulator Systems*", IEEE Transactions on Robotics and Automation, pp 400-410, vol 9, n.4, August 1993

[4] KANG S.B.; Ikeuchi K..;*"Toward Automatic Robot Instruction from Perception-Recognizing a Grasp from Observation"*, IEEE Transactions on Robotics and Automation, pp432-443, vol 9, n.4, August 1993

[5] LOUREIRO, J.A.N.; *"Montagem robotizada-desenvolvimento e integração de CTFMR num manipulador robótico PUMA 560 C";* MSc thesis, Universidade do Minho, Guimarães, Portugal, Oct. 97

[6] PUGH,A.; "*Robot Sensors -Vol2-Tactile and Non-Vision / International Trends in Manufacturing Technology*", IFS Publications Ltd, UK, 1986

[7] QIAO H.; Dalay, B.S.; Parkin R.M.; "*Precise robotic chamferless peg-hole insertion operation without force feedback and remote centre compliance (RCC)* ", ImechE, vol 208, pp 89-104, part C: Journal of Mechanical Engineering Science, 1994

[8] SUGIMOTO,N.; Futsuhara, K.; "*Safety and Technology of safety confirmation type-coordination with robots /Trends in Ergonomics & Human Factors* ", Elsevier Science Publishers, 1987

[9] TUNG, C.P.; Kak, A.C.; "*Automatic learning of assembly tasks using a DataGlove system*", IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots, Pittsburgh, PA, USA, Aug. 1995

[10] WOHLKE, G.; "*The Karlsruhe dextrous hand: grasp planning, programming and real-time control*", IROS`94 IEEE/RSJ/GI International Conference on Intelligent Robots and Systems. Advanced Robotic Systems and the Real World, Munich, Germany, Sep. 1994

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronical design of underwater sensor/actuator robots for cooperative task execution

Pekka Appelqvist, Mika Vainio, Aarne Halme

Automation Technology Laboratory, Helsinki University of Technology,
P.O. Box 3000, FIN-02015 HUT, Finland
http://www.automation.hut.fi

e-mail: pekka.appelqvist@hut.fi
tel: +358-9-4513311
fax: +358-9-4513308

Mechatronical description of an autonomous sensor/actuator robot SUBMAR is presented. The robot is designed to be used in internal monitoring of various liquid processes. Several similar robots equipped with intelligent control architectures and communication skills operate together as a robot society. Test setup to evaluate cooperative task execution with different strategies in 3D process environment is also introduced.

1. INTRODUCTION

Unlike conventional sensors and actuators in various processes, the sensor/actuator robots proposed in this paper are designed to utilize their ability to move along flow in liquid process environments. As a result, a mobile sensor system can cover the whole process volume and provide more complete view to the process than fixed sensors. Furthermore, in some cases it is also possible to influence the process in limited scale with onboard actuators.

To be able to study these issues and test solutions in practice, a robot prototype SUBMAR (Smart Underwater Ball for Measurement and Actuation Routines) has been designed. SUBMAR is a ball-shaped robot with a diameter of 11 cm. The main functional characteristics include semi-active, i.e. vertical, motion capabilities, measurements with different sensors, wireless communication, and multi-purpose tank-actuators.

The mechatronical structure in sensor/actuator robots can be kept relatively simple, if the idea of exact controlling of the robots can be abandoned. This saves a great deal of energy, the most critical resource of all autonomous devices. Instead, obvious uncertainty and stochastic behaviour of the robots due to their incomplete motion capabilities can be compensated by using a number of cooperative robots, see Figure 1.

Several prototype versions of SUBMARs have been constructed. Design principles, mechatronical considerations, and construction of earlier prototypes are described in [1] and [2]. The robot introduced in

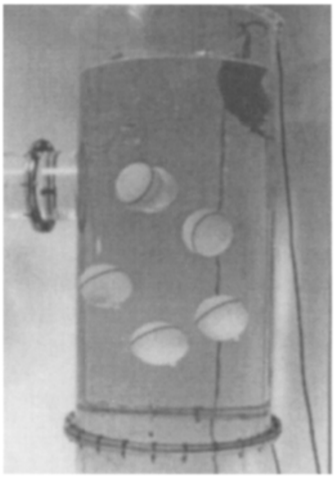


Figure 1. SUBMAR society in test environment

this paper represents the latest version of the third generation prototype. Based on this design, a homogenous society of 10 robots has been produced.

2. COOPERATIVE ROBOT SOCIETY

Distributed multi-robotic systems and utilization of agents' "societal" skills for collective task execution have gained increased research interest. The activity in this area has produced a wide range of approaches to solve the mutual coordination needed in cooperative robot systems, as an example, see [3], and [4].

The concept of *robot society*, introduced originally in [5], is a generic technical concept to describe a group of autonomous mobile robots operating together towards a common goal. Its inspiration comes from Nature, from social animals, like ants and bees. Useful features in the concept of robot society are *flexibility, fault tolerance*, and *simplicity*.

The inbuilt control structure and open communication protocol combined with homogenous society members allow high level of redundancy and flexibility. Volume, i.e. operational efficiency, of the society can be adjusted simply by increasing or decreasing the number of robot members without any reconfiguration. In addition, collective use of several robots enable relatively simple structures for a single robot member.

If one tries moreover to simplify this concept applied to SUBMAR robots, thinking can lead to the world of bacteria. The analogy to bacteria is obvious and intentional, thus the sensor/actuator society can also be called as "bacterium robot society".

3. MECHATRONICAL DESIGN

3.1 Mechanical design

Ball-shaped plastic casing forms mechanically durable structure for SUBMAR. The casing is made of PA6 nylon, which gives good temperature, pressure, shock, and chemical resistant characteristics. Outer diameter is 10.8 cm, having a wall-thickness of 0.6 cm. Different sensors, actuators and electronics are attached into this casing. During sea trials, the construction proved to be fully water resistant at least down to 15 meters.

SUBMARs are equipped with two multi-purpose tank-actuators. They can be used as (1) *Diving tanks* to perform vertical motion and automatic balancing of the specific weight according to environmental conditions, (2) *Chemical tank* to transport reagent to be released at certain location in the process, (3) *Storage for samples* taken from the process.

A tank-actuator consist of a plastic chamber, driving motor, gearbox, screw-like piston-rod, and sealed stainless steel piston. The used motor-gear combination is Maxon 0.5 W DC-motor united to 200:1 reduction gearbox. This will provide enough torque against high outer pressure with moderate power consumption. The construction of tank-actuators is illustrated in Figure 2.

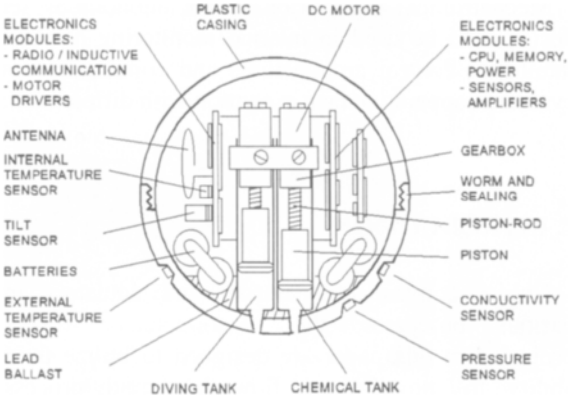


Figure 2. Cross section view of SUBMAR

Compared to the total volume of the robot, approximately 640 ml, the volume of each tank-actuator, 4 ml, is very small. For this reason, the specific weight of the robot has to be initially set relatively close to the surrounding liquid. On the other hand, when the tank is used as a sample extractor for off-line analysis devices, even 1 ml is enough to serve spectral analysators.

3.2 Electronics

While designing electronics for SUBMAR, the main aims have been small size, low power consumption, and modularity. The electronics is based on an efficient 16-bit microcontroller core, as shown in simplified block diagram in Figure 3.

Physically, the electronics is divided into five small cards: CPU, power, amplifier, motor driver, and communication modules. Additionally, actual sensor elements and two battery packs are needed.

Surface mounted components are installed on both sides of the circuit boards to save space. When connected together these cards form a solid structure which can be inserted around the tank-actuators into the casing.

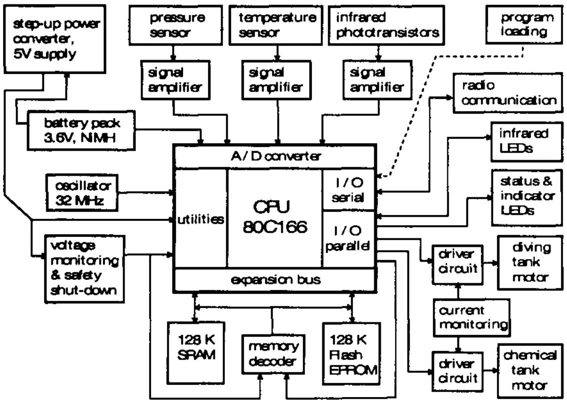

step-up power converter, 5V supply
pressure sensor
temperature sensor
infrared phototransistors
program loading
signal amplifier
signal amplifier
signal amplifier
radio communication
battery pack 3.6V, NMH
A/D converter
infrared LEDs
oscillator 32 MHz
utilities
CPU 80C166
I/O serial
I/O parallel
status & indicator LEDs
voltage monitoring & safety shut-down
expansion bus
driver circuit
diving tank motor
current monitoring
128 K SRAM
memory decoder
128 K Flash EPROM
driver circuit
chemical tank motor

Figure 3. Electronics block diagram

CPU module contains the 16-bit Siemens 80C166 microcontroller, which is operated at 32 MHz clock frequency instead of full 40 MHz to decrease power consumption. As memory circuits, there are 128 Kb of Flash EEPROM for program code and another 128 Kb of SRAM for data storage. I/O-functions, A/D-conversions, and serial communication channels are provided straight from the processor.

Power module has gone under the latest modifications. The previous two separate power units were replaced by one common 5 V supply for all the components. At the same time the input voltage for this step-up converter was dropped from 4.8 V to 3.6 V. This allows to fit two parallel sets of battery packs into the casing. Each pack has three AAU-size NiMH cells, giving a total capacity of 2.7 Ah at 3.6 V. Compared to the previous solution, the battery capacity is increased by 60 %. In operation with normal processor load, sensors, actuators, and communication in use, the average power consumption for SUBMAR is approximately 1.5 W. This allows 6-7 hrs. of operation, depending on the task.

Amplifier module has the needed amplifiers and filters for different type of sensors. Absolute pressure sensor, internal and external temperature sensors, conductivity sensor, as well as an emulated sensors system (see chapter 4.1) for laboratory experiments can be used. Other possible sensors include, for example, pH, dissolved oxygen, turbidity, acceleration.

Motor driver module consists of small MOSFET H-bridge drivers for the tank-actuators. They are monitored by current sensitive detector circuits, which are used to detect the limits of tanks. Otherwise, the tank-actuator pistons are driven only with respect to time, or with feedback from pressure (i.e. depth) sensor.

Communication module uses Radiometrix's BIM half-duplex radio transceiver with a simple whip antenna. The current UHF frequency of 433 MHz is far too high to be used in real world applications in conductive liquids. However, it provides reliable and power effective connections for the operator interface, as well as for robot-to-robot communication for laboratory experiments in fresh water. In these conditions, maximum underwater communication range is about 2 meters. All the electronics modules and other components needed in SUBMAR are presented in Figure 4.

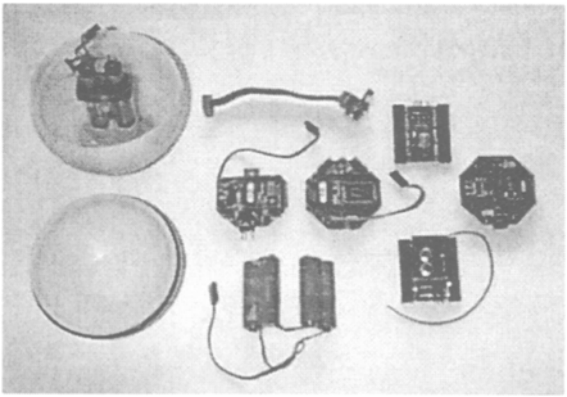


Figure 4. SUBMAR electronics disassembled

3.3 Software concept

The intelligence and adaptivity in cooperative multi-robot system derive from the control architectures of its autonomous members. In the case of SUBMARs, the interrupt-based real-time control architecture is formulated as a generic, hierarchical three-layer Robot Society -model. See [6] and [7] for more comprehensive description of the control architecture. The layers are called *behavioural, task,* and *cooperative layers*.

(1) Behavioural layer takes care of the robot's survival and operational actions. It receives information from sensors, other robots, and operator.

The operational states in this layer can include such as *recover* once the robot gets stucked in some location, *explore* by taking measurements and logging the data to its memory, or *load* to recharge the batteries.

(2) Task layer defines how the robot should proceed when executing its task. The states of the task layer are called *strategies*. Robots optimize their work by choosing the best rewarding strategy.

(3) Cooperative layer is needed to describe *groups* in the society performing simultaneous multi-tasking missions. Robots executing the same task form a group.

Depending on the specified task, the robot has a finite number of possible operational states in each layer. To cope with countless different, often contradictory, requests during operation the robot is allowed to be only in one state at a time. State transitions can be formulated as transition functions. This means that the functioning of each robot can be presented with Finite State Automata. See [8] as an example of similar kind of approach.

3.4 Test environment and navigation

As a laboratory test environment a special demo process has been used. This fully transparent process environment consists of several types and shapes of typical process parts, as seen in Figure 5. The total volume of 700 liters is filled with water, which is circulated with a jet-flow pump. Other instrumentation include a heating resistor, several temperature sensors, an ultrasonic flow-speed meter, and the operator's communication station as user interface to the SUBMAR society.

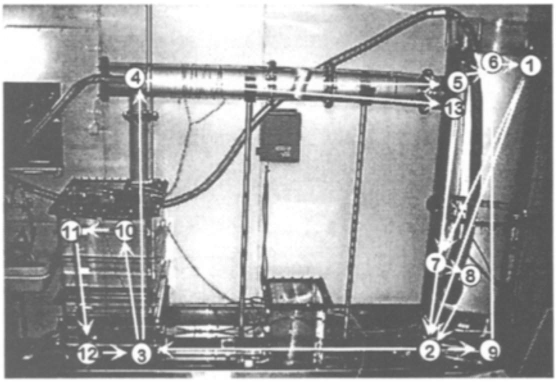


Figure 5. Demo process and the locations found by the feature based positioning system

Positioning and navigation in the demo process environment utilizes so called *feature based positioning system*, which uses an accurate pressure sensor as a primary data source. The algorithms needed in this adaptive method are discussed in [9]. Locations found by this positioning system are drawn also on Figure 5. Based on this map, which is actually a directed graph, a robot can navigate in its environment by following the transitions (links) between "features" in the pressure profile (nodes).

4. COOPERATIVE TASK EXECUTION

4.1 Task definition

In order to study cooperative task execution with the robots, an *explore and exploit* -type of task was defined. The idea is to search and destroy distributed targets, i.e. microbial algae growth spots, in a closed demo process environment. Targets have a dynamic behavior; they grow and get stronger according to a general growth model. If a target is not completely removed, it will continue its growth continuously. The interesting problem is to find an optimal strategy for the robots to remove all the spots as efficiently as possible.

So far, this problem has been examined as a simulator study only, see [7]. By using an emulated sensor system, described in Figure 6., the same studies can be carried out now with real robots in the demo process environment.

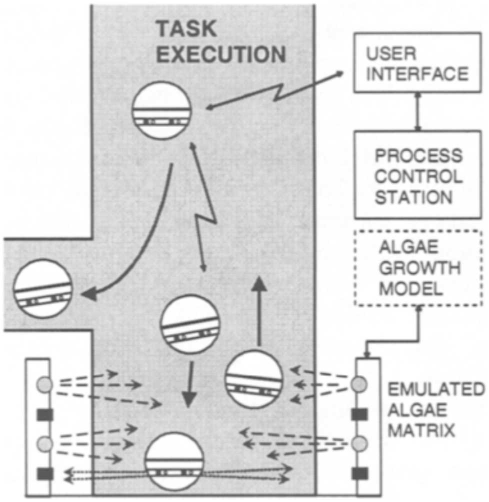


Figure 6. Demotask setup

In the emulated sensor system the robots have IR-photodiodes as their sensors, while lighting by IR-LEDs emulates spreading of poison to be used in algae removal. Respectively, the algae growth spots in the process are presented by infrared LED and photodiode matrix spots. Calculation of algae growth is done as a separate part in the automation system, and driven via control station's I/O.

4.2 Communication and user interface

Communication in the society can be divided into two categories; *between operator and robots*, as well as *inter-robot* communication. Status messages and the measurement data are sent from robot to the operator, while robots can receive higher-level control commands for further operation. Members carry and pass on short messages, distributing information for the whole society.

Since all the data exchange is carried out under a common frequency with half-duplex radio modules, some sort of protocol is required to prevent overloading of the communication network. It is evident that some part of the information will be lost in any case. However, if the communication structure is well designed, some loss will not harm the functioning of the society, but in the worst case only some actions will be delayed.

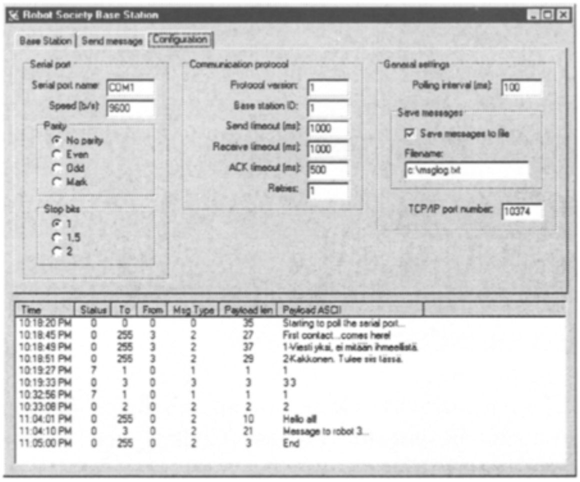
As a solution, a CSMA/CD -type of protocol is being implemented to SUBMARs to enable messaging with minimal loss of information. The protocol frame is presented in Figure 7.

Item	Bytes	Explanation
Message title :		
STX	1	Start byte
Version	1	Version number
Receiver ID	1	Receiver / Broadcast
Sender ID	1	Sender
Message type	1	CMD, INFO, ACK,...
Message ID	1	Used for ACKs
Data length	2	Length of payload
Title CRC	2	Title checksum
Payload CRC	2	Payload checksum
Payload :		
Data max.	65520	Actual message

Figure 7. Message frame for communication

There are some inbuilt features in the protocol to support effective communication: messages can be sent only to certain members, or broadcast to all in the range. In addition, the most important messages can be sent with acknowledge request, to ensure that the message was received correctly. The number of automatic retries can be set, as well as timeouts before them. CRC checksums are calculated separately for the message title and for actual message, payload. To minimize the amount of transferred data, only a collection of fixed messages are used. The protocol does not determine the format of messages.

User interface allows the operator to control the society and get on-line information from the robots. It features protocol settings, different types of sendings, and logging of received data to files. Each robot has its own log file, which allows detailed performance analysis for the task execution evaluation. Figure 8. shows a display from this software; it runs in an NT workstation. The radio module is operated via RS-232 port.



Serial port name: COM1
Speed (b/s): 9600
Parity: ● No parity ○ Even ○ Odd ○ Mark
Stop bits: ● 1 ○ 1.5 ○ 2
Protocol version: 1
Base station ID: 1
Send timeout (ms): 1000
Receive timeout (ms): 1000
ACK timeout (ms): 500
Retries: 1
Polling interval (ms): 100
☑ Save messages to file
Filename: c:\msglog.txt
TCP/IP port number: 10374

```
Time         Status  To   From  Msg Type  Payload len  Payload ASCII
10:18:20 PM  0       0    0     0         35           Starting to poll the serial port...
10:18:45 PM  0       255  3     2         27           First contact...comes here!
10:18:49 PM  0       255  3     2         37           1-Viesti yksi, ei mitään ihmeellistä.
10:18:51 PM  0       255  3     2         29           2-Kakkonen. Tulee siis tässä.
10:19:27 PM  7       1    0     1         1            1
10:19:33 PM  0       3    0     3         3            3 3
10:32:56 PM  7       1    0     1         1            1
10:33:08 PM  0       2    0     2         2            2
11:04:01 PM  0       255  0     2         10           Hello all
11:04:10 PM  0       3    0     2         21           Message to robot 3...
11:05:00 PM  0       255  0     2         3            End
```

Figure 8. View from user interface display

Automatic mission control is a software client for the previously mentioned user interface. It allows pre-programmed mission controls for task execution, i.e. the messages will be automatically send at a given time. TCP/IP connection to the user interface enables also running of the programs from distant location through the internet.

4.3 Performance evaluation

As mentioned before, the described demotask execution has been evaluated so far only with a realistic 3D simulator corresponding to the test setup. The aim of this study was to find out how the performance of the society is affected when the size of the society is varied. The task evaluation was based on three parameters: efficiency in algae removal, elapsed operation time, and number of lost robot members in dead-lock situations.

The obtained results verified that in the particular demo environment an optimal number of the society members could be found. See [7] for more detailed description of the results.

5. CONCLUSIONS

In this paper the mechatronical design of underwater sensor/actuator robots is presented, as well as the communication architecture for cooperative task execution. The emulated sensor system is currently being implemented into SUBMAR robots and the demo process environment to enable the final testing with real robots.

ACKNOWLEDGMENTS

The authors would like to thank Markku Kontio for his excellent effort in the user interface software.

REFERENCES

[1] P. Appelqvist: Design of an Intelligent Sensor /Actuator Robot for Process Control and Robot Society Studies, *Master's Thesis*, Helsinki University of Technology, 1996.

[2] P. Appelqvist, A. Halme, T. Schönberg, M. Vainio, Y. Wang: Designing simple cooperative sensor/actuator robots for liquid process environments, *Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics AIM'97* (CD ROM), Tokyo, Japan, 1997.

[3] *Distributed Autonomous Robotic Systems*, H. Asama, T. Fukuda, T. Arai, I. Endo, Asama, (Eds.), Springer-Verlag, Tokyo, 1994.

[4] *Distributed Autonomous Robotic Systems 2*, H. Asama, T. Fukuda, T. Arai, I. Endo, Asama, (Eds.), Springer-Verlag, Tokyo, 1996.

[5] A. Halme, P. Jakubik, T. Schönberg, M. Vainio: The concept of robot society and its utilization, *Proceedings of the 1993 IEEE International Workshop on Advanced Robotics*, Tsukuba, Japan, pp. 29-35, 1993.

[6] M. Vainio, P. Appelqvist, T. Schönberg, A. Halme: Group behavior of a mobile underwater robot society destroying distributed targets in a closed process environment, *Proceedings of the 3rd IFAC Symposium on Intelligent Autonomous Vehicles IAV'98*, Madrid, Spain, pp. 112-117, 1998.

[7] M. Vainio, P. Appelqvist, A. Halme: Generic Control Architecture for a Cooperative Robot System, *forthcoming in IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'98*, Oct. 12.-16., Victoria, B. C., Canada, 1998.

[8] D. C. MacKenzie, R. C. Arkin, J. M. Cameron, Multiagent mission specification and execution, *Autonomous Robots*, 4(1), pp. 29-52, 1997.

[9] A. Halme, P. Appelqvist, P. Jakubik, P. Kähkönen, T. Schönberg, M. Vainio: Bacterium Robot Society - A Biologically Inspired Multi-Agent Concept for Internal Monitoring and Controlling of Processes, *Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Osaka, Japan, pp. 1707-1714, 1996.

SUBMAR® is the registered trademark of ISS Ltd.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Neural Network Adaptive Controller for DD Robot

R. Safaric[a], K. Jezernik[a], M Kupljen[a], R. M. Parkin[b]

[a]Faculty of Electrical Engineering and Computer Science, University of Maribor,
Smetanova 17, 2000 Maribor, Slovenia

[b]Mechatronics Research Group, Loughborough University,
Loughborough, Leicestershire, LE11 3TU, United Kingdom

A comparision between a Lyapunov based neural network continiuous sliding-mode conroller (NNSMC) and a computer torque based neural network controller (NNCTC) for a direct-drive robot mechanism is shown in the paper. Controllers were verified on a direct-drive 3.D.O.F. PUMA mechanism. Both neural network controllers were successfully tested for trajectory tracking control tasks with respect to two criteria:
-properties of the proposed control algorithm (high speed cyclic movement) convergence and
-adaptation capability of the algorithm for sudden changes in the manipulator dynamics (load).

1. INTRODUCTION

High precision and high speed trajectory tracking direct-drive (DD) robotic manipulators have become increasingly important in the field of flexible automation. However, the derived dynamic mathematical model equations of the DD robot mechanism show its high non-linearity. Therefore structured and/or unstructured uncertainties [1] in case of the DD robot mechanism are even more important than in the geared robot mechanism where influences of the mechanism on the AC-motors are decreased due to the high gear ratios in gear boxes.

The need of meeting demanding control requirements in increasingly complex dynamical control systems under significant structured and/or unstructured uncertainties makes neural networks attractive because of their ability to learn by experience and generalize the acquired knowledge about system dynamics. Through the learning process, the model uncertainties are eliminated, and thus neural networks (NN) provide the necessary compensation in the robot controller.

A number of publications dealing with the topic of the trajectory tracking neural network controllers based on the computed torque method mostly on the level of simulations [2],[3],[4] and [1] etc. have been published in recent years. In sources [2] and [3] they tried to replace the estimated model of the real mechanism with two neural networks. The disadvantage of this method is that it requires generalized learning [3] in addition to specialized learning or a time-consuming convergence of neural network learning [2] if generalized learning is not implemented. In order to speed up the convergence, without generalized learning, the source [4] retained the complete compensator based on the computed torque method and added a neural network approximating an unstructured uncertainty, which would not be compensated by the computed torque method itself and would introduce an error into the control system if used with this method. The disadvantage of the method described in the source [4] is that the exact robot model has to be known. In order to diminish the drawbacks of all the mentioned methods, we tried to reach a compromise between the methods of the sources [4] and [2]. Thus we decided to use one neural network because of a higher convergence speed with as little preliminary knowledge on the estimated mechanism model as possible. In both cases we used robust control schemes: a continiuous sliding mode control [6] and a computed torque method [2] which were used for avoiding problems about a start of neural network learning from randomly choosen initial weights.

2. CONTROL METHODS

2.1. Three D.O.F. direct-drive PUMA like robot mechanism model

The mathematical model of the direct drive 3 D.O.F. PUMA mechanism is presented with equation 1 [8]. Data of the robot mechanism are shown on WEB site:

http : //www.ro.feri.uni – mb.si/projekti/4_dd_drive.html

$$T = M.\ddot{\theta} + h + G_f + T_n \qquad (1)$$

where T, h, G_f and T_n are column vectors of the 3 x 1 dimension, M is a matrix of the 3 x 3 dimension, and $\theta = [\theta_1 \ \theta_2 \ \theta_3]^T$ is a column vector of the 3 x 1 dimension of all three joint positions.

2.2. Neural network continiuous sliding-mode controller

The Lyapunov based theoretical dvelopment of the NNSMC is shown in [10]. Equation 1 has been transformed as follows:

$$\dot{x} = f + \Delta B.u + \tilde{B}.u + d \qquad (2)$$

where

$$x = [\theta_1\theta_2\theta_3\dot{\theta}_1\dot{\theta}_2\dot{\theta}_3]^T; \qquad \dot{x} = [\dot{\theta}_1\dot{\theta}_2\dot{\theta}_3\ddot{\theta}_1\ddot{\theta}_2\ddot{\theta}_3]^T \qquad (3)$$

$$\mathrm{f} = \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \\ -\hat{M}^{-1}.[\hat{h} + \hat{G}_f + \hat{F} + \hat{T}_n] \end{bmatrix}; \mathbf{B} = \begin{bmatrix} 0 \ 0 \ 0 \\ 0 \ 0 \ 0 \\ 0 \ 0 \ 0 \\ \hat{M}^{-1} \end{bmatrix} \qquad (4)$$

where $\hat{M}$, $\hat{h}$, $\hat{G}_f$ and $\hat{T}_n$ are estimated values of M, h, G_f and T_n. The matrix $\hat{M}$ has nominal or average parameters. It means that all nine parameters of the matrix $\hat{M}$ are constant during all robot tasks. The dimensions of the vector f are 6 x 1, and the dimensions of the matrix $\tilde{B}$ and ΔB are 6 x 3.

The control law u is described by the following equation:

$$u = -(G.\tilde{B})^{-1}.[G.(N - \dot{x}_r) + D.\sigma] \qquad (5)$$

where

$$G = \begin{bmatrix} K_{p1} & 0 & 0 & K_{v1} & 0 & 0 \\ 0 & K_{p2} & 0 & 0 & K_{v2} & 0 \\ 0 & 0 & K_{p3} & 0 & 0 & K_{v3} \end{bmatrix}; D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \qquad (6)$$

The coefficients of the matrices G and D are selected in such a manner that the most rapid convergence of the neural network learning algorithm is made possible.

$$x_r = [\theta_{1r}\theta_{2r}\theta_{3r}\dot{\theta}_{1r}\dot{\theta}_{2r}\dot{\theta}_{3r}]^T; \qquad (7)$$

$$\sigma = G.(x - x_r); \qquad \dot{\sigma} = G.(\dot{x} - \dot{x}_r) \qquad (8)$$

The column vector N is of the 6 x 1 dimension and represents the outputs of the multilayer perceptron neural network [7] o_i $(i = 1...6)$. The dimension of the control column vector u is 3 x 1. The learning procedure for all weights of the neural network output layer is:

$$\begin{aligned} \Delta w_{1j} &= \eta.[K_{p1} \quad 0 \quad 0].(D.\sigma + \dot{\sigma}).g'(net_1).o_j \\ \Delta w_{2j} &= \eta.[0 \quad K_{p2} \quad 0].(D.\sigma + \dot{\sigma}).g'(net_2).o_j \\ \Delta w_{3j} &= \eta.[0 \quad 0 \quad K_{p3}].(D.\sigma + \dot{\sigma}).g'(net_3).o_j \\ \Delta w_{4j} &= \eta.[K_{v1} \quad 0 \quad 0].(D.\sigma + \dot{\sigma}).g'(net_4).o_j \\ \Delta w_{5j} &= \eta.[0 \quad K_{v2} \quad 0].(D.\sigma + \dot{\sigma}).g'(net_5).o_j \\ \Delta w_{6j} &= \eta.[0 \quad 0 \quad K_{v3}].(D.\sigma + \dot{\sigma}).g'(net_6).o_j \end{aligned} \qquad (9)$$

where

$$net_i = \sum_j (w_{ij} o_j) + b_i; \;\; NN(\theta, \dot{\theta}) = o_i = g(net_i) \qquad (10)$$

$$g(net) = 1 - 2/(1 + e^{-net}) \qquad (11)$$

where $j = 1...60$, $i = 1...6$, $l = 1...9$ and $g'(*)$ is first derivative of sigmoidal function (equation 11). Nine neural network inputs were used: three actual positions, three actual velocities and three differences between desired and actual positions in the joint space. The control scheme of described neural network sliding mode controller is presented in figure 1.

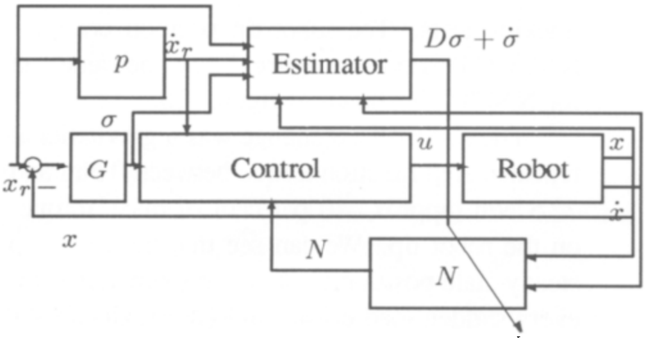

p
$\dot{x}_r$
Estimator
$D\sigma + \dot{\sigma}$
σ
G
Control
u
Robot
x
$\dot{x}$
x_r
x
N
N

Figure 1: Control scheme of NNSMC

2.3. Neural network computed torque method controller

Figure 2 and equation 12 show the control law for the NNCTC. We use a NN for aproximation of structured and/or unstructured uncertainties instead of computed torques for $\hat{h} + \hat{G}_f + \hat{F} + \hat{T}_n$ [2].

$$T = \hat{M}(\theta).u + NN(\theta, \dot{\theta}) \qquad (12)$$

where

$$u = \ddot{\theta}_{Ref} + K_v.\dot{E} + K_p.E; \;\; E = \theta_{Ref} - \theta \qquad (13)$$

We use so called CMAC NN for the NNCTC. A schematic of the CMAC NN [11] is shown in figure 3. Algorithm for learning method is described by equation 14.

$$\Delta w = lr.(T - A)/C \qquad (14)$$

where T is the desired output of the CMAC NN, A is the actual output of CMAC NN, lr is the learning rate, C presents the number of layers in the CMAC NN and Δw is the calculated weight change in one sampling period ($T_s = 2ms$).

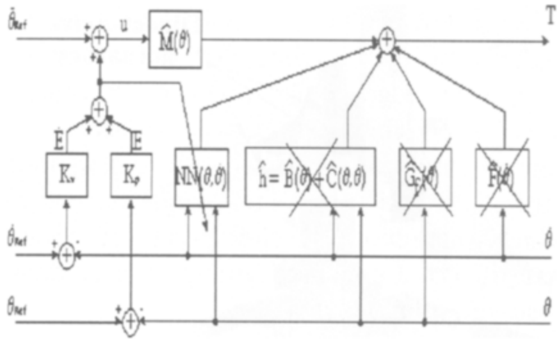

$\ddot{\theta}_{Ref}$
u
$\hat{M}(\theta)$
T
$\dot{E}$
E
K_v
K_p
$NN(\theta,\dot{\theta})$
$\hat{h}=\hat{B}(\theta)+\hat{C}(\theta,\dot{\theta})$
$\hat{G}_f(\theta)$
$\hat{F}(\dot{\theta})$
$\dot{\theta}_{Ref}$
$\dot{\theta}$
θ_{Ref}
θ

Figure 2: Control scheme of NNCTC

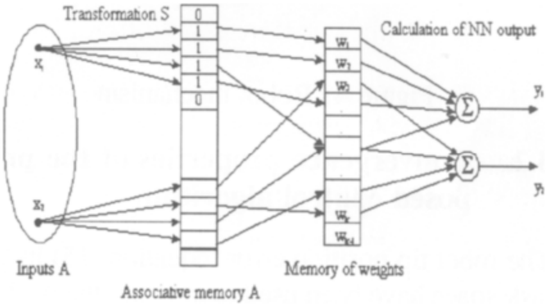

Transformation S
Calculation of NN output
x_1
x_2
y_1
y_2
Inputs A
Associative memory A
Memory of weights

Figure 3: CMAC Neural Network

3. APPLICATION RESULTS

3.1. The realization of the robot mechanism

Dynaserv AC-motors produced by the Yokogawa company with maximal torque of 220Nm, 160Nm and 60 Nm and nominal angular velocity of 2 rotations per second are used. The robot's body is made of steel, whereas the connection elements are made of aluminum alloy, welded and fixed to the motors. A wrist or a welding device can be added to the robot tip. Figure 4 shows robot mechanism. More precise explanation of the robot mechanism is made in [9]. As the robot mechanism should be used for welding

or manipulation, several tasks were tested:
1. high speed robot tip trajectory tracking (over 1 m/sec) and
2. robustness to load torque changes.

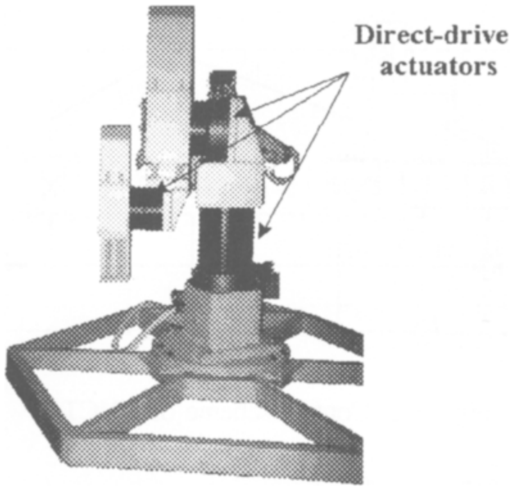


Figure 4: Robot mechanism

3.2. Convergence properties of the proposed control algorithm

The robot tip position error (equation 15) in the task space have been used to measure the quality of the robot trajectory tracking and robustness to load torque changes:

$$E_p = [(X_{di}-X_i)^2+(Y_{di}-Y_i)^2+(Z_{di}-Z_i)^2]^{1/2} \tag{15}$$

where X_{di}, Y_{di}, Z_{di} are reference trajectories in the i-th sampling time in the task space. X_i, Y_i, Z_i are the actual trajectories in the i-th sampling time in the task space.

The test trajectory for the high speed cyclic PTP movements are shown in figures 5 to 8. Figure 5 shows the robot tip cyclic reference trajectory in the three dimensional (X-Y-Z) task space. Figure 6 shows the robot tip reference speed and acceleration in the task space. Figure 7 shows the difference between the reference and actual robot tip position for the NNSMC in the task space and figure 8 shows the same difference for the same conditions for the NNCTC. It is evident from figures 5 to 8 that convergence is complete after twelve learning cycles (the learning cycle lasts 4.8 sec.), for which the initial weights of the neural network were randomly chosen; the robot tip position error in the task space is less than 2 cm for the NNSMC amd less than 1.5cm for the NNCTC for high speed (1m/sec) cyclic PTP movement.

3.3. Adaptation capability for sudden load changes

The test of sudden load changes in stationary position has the same initial conditions for both tested controllers as the high-speed trajectory tracking tests. The robot tip position error for NNSMC is shown in figure 9 and the same test for NNCTC is shown in figure 10.

The sudden load change was a gravitational repeated step function torque between 0 Nm and 22.5 Nm (approx. 40 per cent of max. torque) on the robot tip. We can see that the robot tip steady-state position error is approximately 0 for every sudden load change after approximately 1 sec. transient time in the case of the NNSMC (figure 9). It is also evident from the same figure that error peaks are lower than 5 mm. Figure 10 shows the robot tip position error for the NNCTC during only two steps of load changes from 0 Nm to 22.5 Nm (at the time 22 sec) and vice versa(at the time 26 sec). We can observe from the figure that the error peaks are higher (1-2 cm) than in the case of the NNSMC. It is also evident that the NNCTC response is almost unstable in the presence of 22.5 Nm step disturbance on the robot tip. We can also see in the same figure that the transient time after change of the step load disturbance from 22.5 Nm to 0 Nm is about 12 sec which is much longer than in the case of the NNSMC.

We also added two more figures with responses to the same disturbance test for two different non-neural network controllers. Figure 11 shows the robot tip position error for the continuous sliding mode controller (CSMC) with PI disturbance observer during the load disturbance

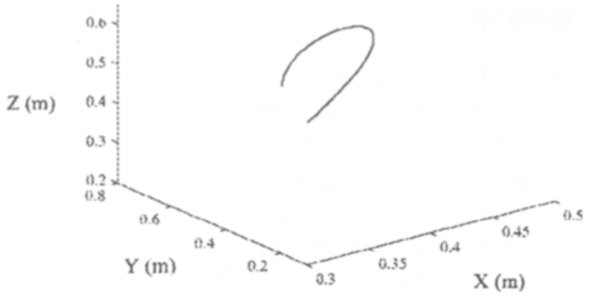


Figure 5: Robot tip reference position

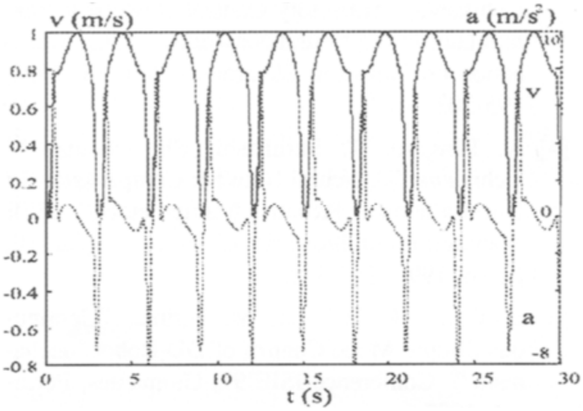


Figure 6: Robot ref. velocity and acceleration

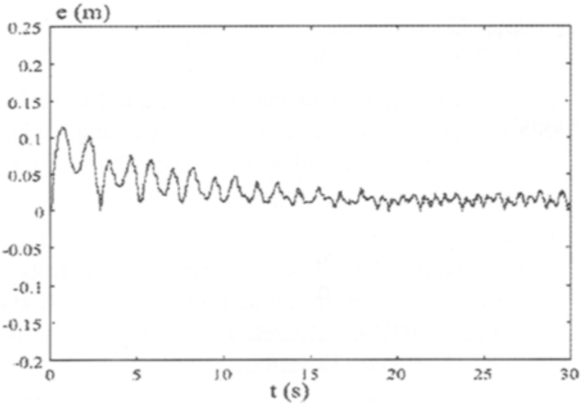


Figure 7: Robot tip trajectory tracking position error for NNSMC

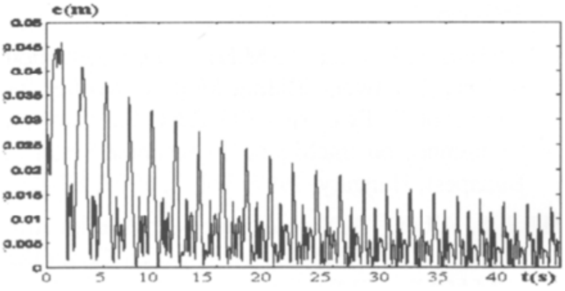


Figure 8: Robot tip trajectory tracking position error for NNCTC

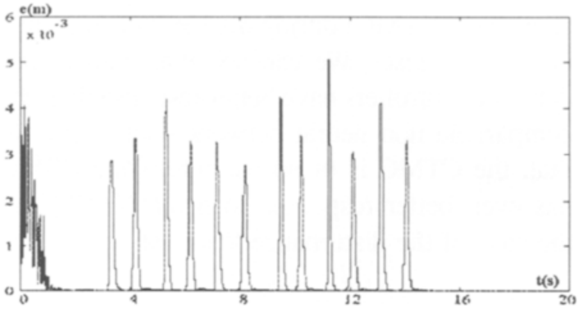


Figure 9: Robot tip position error for NNSMC during load changes

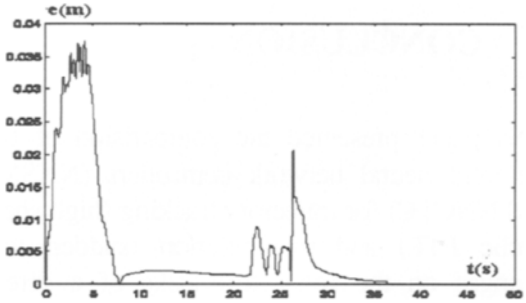


Figure 10: Robot tip position error for NNCTC during load changes

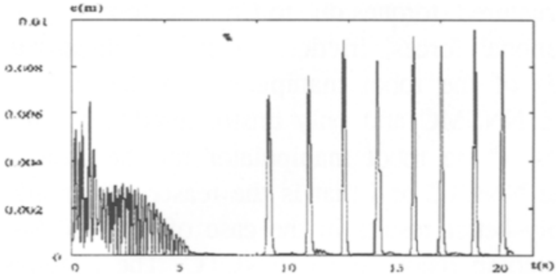


Figure 11: Robot tip position error for CSMC with PI-estimator during load changes

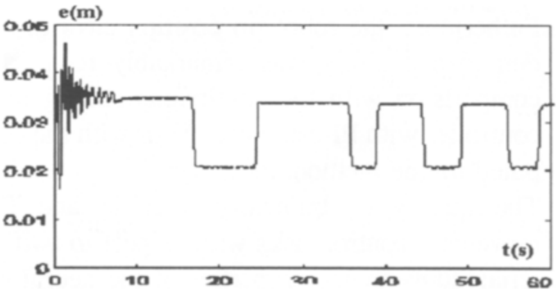


Figure 12: Robot tip position error for CTMC during load changes

changes whilst figure 12 shows the robot tip position error for the computed torque method controller (CTMC) during the mentioned disturbance load test. We can see that both neural network controllers have better response than the comparable non neural network controllers. In fact, the CTMC is of no use whilst the CSMC has even better responses than the NNCTC in the case of the disturbance load test.

4. CONCLUSION

This paper presented the comparision of two different neural network controllers (NNSMC and NNCTC) for trajectory tracking (high speed cyclic PTP) and manipulation (sudden load changes on the robot tip) tasks of a direct-drive three D.O.F. PUMA like robot manipulator. The neural network was used to compensate structured (the inertia matrix) and unstructured (torques due to Coriollis forces, gravitational forces, friction forces ...) uncertainties of the robot manipulator in the case of the NNSMC and only unstructured uncertainties of the robot manipulator for the case of the NNCTC and that is the reason of remarkably better results in the case of the NNSMC in comparision to the NNCTC. The adaptive and self-improving capability of the both neural network controllers to structured and unstructured effects (sudden load changes) was shown. Futhermore, the robot tip position error at sudden load changes was remarkably reduced in comparision with the continuous sliding mode controller with PI-estimator [5] or with the computed torque method controller.

The quality of trajectory tracking and PTP movement control tasks with respect to two criteria: convergence properties of the neural network control algorithm and adaptation capability of the control algorithm to sudden changes in the manipulator dynamics were shown.

References

[1] S. Khemaissia, A. S. Morris, "Neuro-adaptive control of robotic manipulators," Robotica, vol.11, 1993, 465-473.

[2] R. Safaric, K. Jezernik, "Trajectory Tracking Neural Network Controller for a Robot Mechanism and Lyapunov Theory of Stability," Conference Proceedings IROS'94, Munich, 1994.

[3] T. Ozaki, T. Suzuki, T. Furuhashi, S. Okuma, Y. Uchikawa, "Trajectory Control of Robotic Manipulator Using Neural Networks," IEEE Transaction on Industrial Electronics, Vol. 38, 1991, 195-202.

[4] A. Ishiguro, T. Furuhashi, Sh. Okuma, Y. Uchikawa, "A Neural Network Compensator for Uncertainties of Robotics Manipulators," IEEE Transactions on Industrial Electronics, VOL. 39, No. 6, 1992, 555-570.

[5] B. Curk, K. Jezernik, M. Terbuc, "Continuous Sliding Mode Control of DD Robot Mechanism," Conference ISIE'97, Guimaraes, Portugal, 1997.

[6] A.Sabanovic, "Sliding Mode in Robotic Manipulator Control Systems," Electrotechnical review, 60(2-3), 1993 99-107.

[7] A. Cichocki, R. Unbehauen, Neural Networks for Optimization and Signal Processing, John Wiley and Sons, Chichester - New York- Brisbane - Toronto - Singapore, 1993.

[8] M. Terbuc, M. Rodic, K. Jezernik, "Direktno gnani robotski sistem(Direct Drive Robot System)," Zbornik cetrte Elektrotehniske konference ERK'95, Zvezek B, Slovenska sekcija IEEE, Portoroz, Slovenia, 1995, 187-190.

[9] K. Hornik, M. Stinchombe, H. Whiteat, "Multilayer Feedforward Networks are Universal Aproximators," Neural Networks, Vol. 2, 1989, 359-366.

[10] R.Safaric, K.Jezernik, M.Pec, "Implementation of Neural Network Sliding Mode Controller for DD robot.," Proc. of 1997 IEEE International Conference on Intelligent Engineering Systems, Budapest, Hungary, 1997, Vol. 2, 1997, 83-88.

[11] J. S. Albus, "A New Approach to Manipulator Control: The Cerebellar Articulation Controller (CMAC)," ASME, J. Dynamics Systems, Measurment and Control, Vol. 97, No.3 1975.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Adaptive Tracking Error Controller for Rigid-Link Flexible-Joint Robot Manipulators

C. J. Tsaprounis, N. A. Aspragathos
University of Patras

Department of Mechanical and Aeronautical Engineering

Patras 26500 Greece

Email address: tsaproun@mech.upatras.gr, asprag@mech.upatras.gr

This paper presents an adaptive control scheme for flexible joint robot manipulators. The asymptotic stability of the control law is insured using Lyapunov Theory. The feedback to the control law is the measurement of the joint torques and the link positions and velocities. The characteristics of both flexible and rigid subsystems are assumed unknown. An adaptive control law estimates the values of the unknown parameters. The performance of the proposed control method is tested using a simulated planar robot with two rotational degrees of freedom.

1. INTRODUCTION

Robot controller task is a representative mechatronical task. The design of controllers for rigid arms has reached a high level of sophistication despite the nonlinearity, and the strong interactivity of the multivariable dynamic mechanical system of manipulators. However, the development of light, accurate, and fast robots able to interact gently with rigid environments requires the synergetic integration of mechatronical engineering. In order to build robots with these advanced characteristics it is almost obligatory to design flexible links and joints. The flexibility of links and the joints complicates much more the controller design problem. Therefore, this generation of robots can not be designed sequentially as it is the conventional way: first design the mechanisms and then adapt the controller. The designer of the mechanical part it has to take into account the advancements of the intelligent control based on the power of modern computers and to integrate sensors considering the requirements of the new control algorithms.

The assumption of perfect rigidity is ideal for all manipulators. The main source of flexibility is the joint of robot manipulator, and to some extend it is presented in the current industrial robots. Since the robot links are stiffer than the joints, therefore in the present analysis the links are assumed rigid.

Undesirable vibrations and non-negligible position errors are caused by the flexibility of robot joints. In most cases, this intrisinc small position error is regarded as a source of problems, especially when accurate trajectory tracking or high sensitivity to end-effector forces is mandatory. On the other hand, there are cases where compliant joints may become useful in a robotic structure, e.g. as a protection against "hard" contacts during assembly tasks [1].

Several methods have been employed in the design of controllers for flexible joint manipulators. The classical approach is based either on the feedback linearization or on the singular perturbation method. Last years adaptive algorithms are introduce for the control of flexible joint robots.

Rogelio Lozano et al [2] presented an adaptive algorithm for the joint control of flexible joint robots. An adaptive estimator used for the estimation of the unknown parameters of the system (inertia properties stiffness of joints etc.). The singularities in the estimation of the unknown parameters are avoided due to the design of the adaptive law. Their control law requires the double differentiation of inertial and centrifugal matrices of the rigid subsystem and the triple differentiation of the actuator position. This increases the computational cost for the implemented controller, while the way that the three times differentiation is influenced by the existent of noise in the measuring data needs further investigation.

Ser Yong Lim et al [3] presented an adaptive controller where joint velocity measurements are not needed. A set of high pass filters is used for the estimation of velocity errors of links and joints. Based on this set of filters, an adaptive integrator backstepping procedure is used to develop the needed input torque. The presented results show that the algorithm is quite robust and rejects the noise presented in the measured position of the links and actuators.

Mrad and Ahmad [4] presented an adaptive control law based on the feedback of the velocities and positions of robot actuators and links. The characteristics of robot links, joints and actuators are determined through a second order adaptive estimator based on the tracking error of the robot links, therefore the link acceleration is not needed in the control system. From the simulation results it is concluded that the initial values of the estimated parameters are chosen quite close to the real ones, which results to a faster convergence of the tracking error. Chattering effect can be seen in the response graphs, which can be eliminated by using a boundary layer as the authors pointed out.

Song et al [5] presented a robust controller for flexible joint robots executing constraint motion tasks encountering friction forces at the end-effector. A smooth robust compensator was designed to stabilise the flexible subsystem in order to generate the desired torque for the rigid subsystem. In the algorithm the exact values of rigid and flexible subsystem characteristics are not necessary. The presented results from the numerical simulation show a tiny tracking error and a perfect compensation of the friction force.

Pheffer et al [6] designed a controller based on joint torque sensory feedback for PUMA 500 manipulator which was implemented in the third joint having a total stiffness about 297 Nm/rad. The authors claim that the designed controller with joint torque feedback provides a substantial reduction of the effect of friction at the transmission system. From the experimental results it is obvious that the performance of the controller is very good. The implementation of this controller show that joint torque feedback can be used without serious problems related to the resolution of the joint torque sensor and its signal noise.

Zhang et al [7] presented a control algorithm where an observer based on torque feedback is used for friction compensation. From the numerical results it is shown that the performance of this algorithm is better than other algorithms using only position and velocity feedback.

In this paper an adaptive control algorithm for flexible joint manipulators is presented. The algorithm is based on the feedback of joint torques and their derivative and on the signals of link position, velocity and acceleration. The differential equations for the manipulator are re-written in order to introduce the joint torques as a state space variable. It is proved that the system is marginally stable. The use of joint torques in the control algorithm makes easier its implementation and there is no need for any assumption about the stiffness of the joints.

In the following analysis the derivative of the joint torques and the link accelerations are needed, which can be calculated by the control system quite easily [5], [6]. In addition, the desired motion should be at least four times differentiable. An adaptive law estimates the parameters of the links and joints. The global asymptotic stability of the control law is proved via Lyapunov Direct method.

1 DYNAMICS OF FLEXIBLE JOINT ROBOT MANIPULATORS

It is assumed that a revolute flexible joint direct-drive robot is described by the following differential equations [1]:

$$\begin{aligned} & H_r \underline{\ddot{q}} + C_r \underline{\dot{q}} + \underline{g}_r = \underline{\tau}_J \\ & H_m \underline{\ddot{\theta}} + \underline{\tau}_J = \underline{\tau} \\ & \underline{\tau}_J = K_J (\underline{\theta} - \underline{q}) \end{aligned} \qquad (1)$$

where, H_r, $C_r \in R^n \times R^n$ are the inertia and coriolis matrices of the rigid subsystem, the vector $\underline{g}_r \in R^n$ represents the gravity forces for the rigid subsystem, the H_m, $K_J \in R^n \times R^n$ are the inertia matrix of actuators and the joint stiffness matrix, the vectors $\underline{q}$, $\underline{\theta} \in R^n$ are the angular displacements of the links and the actuators, the vectors $\underline{\tau}$, $\underline{\tau}_J \in R^n$ are the torque input and the joint torque respectively, and n is the number of robot DOF.

The robot model described by Eq.(1) represents the class of anthropomorphic manipulators. Furthermore the coupling inertial term is assumed negligible in comparison to the inertia of the drive system.

Since the developed control law is based on joint torque feedback, the equations (1) are rewritten in the following form, where the joint torques are considered as state space variables.

$$\begin{bmatrix} H_r & 0 \\ H_m & H_m K_J^{-1} \end{bmatrix} \begin{bmatrix} \underline{\ddot{q}} \\ \underline{\ddot{\tau}}_J \end{bmatrix} + \begin{bmatrix} C_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \underline{\dot{q}} \\ \underline{\dot{\tau}}_J \end{bmatrix} + \qquad (2)$$

$$+ \begin{bmatrix} 0 & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} \underline{q} \\ \underline{\tau}_J \end{bmatrix} = \begin{bmatrix} -\underline{g}_r \\ \underline{\tau} \end{bmatrix}$$

From Eq.(2) it can be concluded that the system has zero eigenvalues. These are introduced by the rigidity of the links which results to the zero left upper submatrix of the zero-order term of Eq.(2). In addition the term $H_m K_J^{-1}$ can not be neglected when the joint stiffness is small. Therefore, the system is marginally stable and the low stiffness of joints introduces forces related to the second derivative of joints torque.

2 DESIGN OF CONTROL LAW

The proposed control algorithm is based on the following assumptions: The position, velocity and acceleration of the rigid subsystem are available or can be calculated while the joint torque is measured and its first derivative is calculated. Finally, the dynamic characteristics of the links, joints and actuators are not known.

For the flexible subsystem the following sliding surface is formed:

$$\underline{s}_r = \left(\dot{\underline{q}} - \dot{\underline{q}}_d\right) + \Lambda_r\left(\underline{q} - \underline{q}_d\right) \quad (3)$$

where, $\underline{q}_d$ represents the desired trajectory and $\dot{\underline{q}}_d$ represents the desired velocity of the rigid subsystem and Λ_r is a positive defined matrix.

The estimated joint torques are given by the following equation:

$$\begin{aligned}\hat{\underline{\tau}}_{Jd} &= Y_d\left(\underline{q}_d, \dot{\underline{q}}_d, \ddot{\underline{q}}_d\right)\hat{\underline{p}}_r \\ &= \hat{H}_r \ddot{\underline{q}}_d + \hat{C}_r \dot{\underline{q}}_d + \hat{\underline{g}}_r\end{aligned} \quad (4)$$

where $\hat{\underline{\tau}}_{Jd}$ represents the estimated desired joint torque, $Y_d \in R^n \times R^m$ is a regression matrix and $\hat{\underline{p}}_r \in R^m$ is the vector of the estimated constant parameters of the matrices H_r, C_r and of the vector $\underline{g}_r$. In the following analysis the symbol $\wedge$ is used for the estimated value of a quantity. In the same way, the symbol $\sim$ is used for the error between the estimated and the real value of a quantity.

From Eq.(4) is obvious that in order to determine the second derivative of the estimated desired torque the fourth derivative of the desired link position is needed. As it is shown later the second derivative of the estimated desired trajectory is needed to determine the control law.

For the flexible subsystem the following sliding surface is formed:

$$\underline{s}_f = \left(\dot{\underline{\tau}}_J - \dot{\hat{\underline{\tau}}}_{Jd}\right) + \Lambda_r\left(\underline{\tau}_J - \hat{\underline{\tau}}_{Jd}\right) \quad (5)$$

The reference velocity is given by:

$$\nu_f = \dot{\hat{\underline{\tau}}}_{Jd} - \Lambda_r\left(\underline{\tau}_J - \hat{\underline{\tau}}_{Jd}\right) \quad (6)$$

while the proposed control law is given by:

$$\begin{aligned}&If\ \left\|\underline{s}_f\right\|_2 \neq 0 \\ &\tau = Y_{m1}\hat{\underline{p}}_{m1} + Y_{m2}\hat{\underline{p}}_{m2} + \underline{\tau}_J - K_f \underline{s}_f + \quad (7) \\ &+ \frac{\underline{s}_f}{\left\|\underline{s}_f\right\|_2}\underline{s}_r^T\left(Y_\nu \hat{\underline{p}}_r - K_r \underline{s}_r - \underline{\tau}_J\right)\end{aligned}$$

and

$$\begin{aligned}&If\ \left\|\underline{s}_f\right\|_2 = 0 \\ &\tau = Y_{m1}\hat{\underline{p}}_{m1} + Y_{m2}\hat{\underline{p}}_{m2} + Y_\nu \hat{\underline{p}}_r - K_r \underline{s}_r\end{aligned} \quad (8)$$

where

$$\hat{H}_m \hat{K}^{-1}\dot{\underline{\nu}}_f = Y_{m1}\left(\dot{\underline{\nu}}_f\right)\hat{\underline{p}}_{m1} \quad (9)$$

$$\hat{H}_m \ddot{\underline{q}} = Y_{m2}\left(\ddot{\underline{q}}\right)\hat{\underline{p}}_{m2} \quad (10)$$

The matrices $K_f \in R^n \times R^n$ and $K_r \in R^n \times R^n$ are positive definite and fed by the designer of the control system. Based on the linearity property of the dynamic parameters $\hat{\underline{p}}_r$ of the robot model the following equation is true:

$$Y_\nu\left(\underline{\nu}_r, \dot{\underline{\nu}}_r, \underline{q}, \dot{\underline{q}}\right)\hat{\underline{p}}_r = \hat{H}_r \dot{\underline{\nu}}_r + \hat{C}_r \underline{\nu}_r + \hat{\underline{g}}_r \quad (11)$$

The following adaptive laws are selected for the estimation of the unknown parameters:

$$\left.\begin{aligned}&For\ \left\|\underline{s}_f\right\|_2 \neq 0 \\ &\dot{\hat{\underline{p}}}_r^T = -\Gamma_r \underline{s}_r^T Y_\nu,\ \dot{\hat{\underline{p}}}_{m1}^T = -\Gamma_{m1}\underline{s}_f^T Y_{m1} \\ &\dot{\hat{\underline{p}}}_{m2}^T = -\Gamma_{m2}\underline{s}_f^T Y_{m2} \\ &For\ \left\|\underline{s}_f\right\|_2 = 0 \\ &\dot{\hat{\underline{p}}}_r^T = -\Gamma_r \underline{s}_r^T Y_\nu,\ \dot{\hat{\underline{p}}}_{m1}^T = -\Gamma_{m1}\underline{s}_r^T Y_{m1} \\ &\dot{\hat{\underline{p}}}_{m2}^T = -\Gamma_{m2}\underline{s}_r^T Y_{m2}\end{aligned}\right\} (12)$$

where, $\Gamma_r \in R^n \times R^m$, $\Gamma_{m1} \in R^n \times R^n$ and $\Gamma_{m2} \in R^n \times R^n$ are positive definite adaptation matrices. For the stability study of the system along the desired trajectory the following Lyapunov candidate function is defined:

$$\begin{aligned}V(t) &= \frac{1}{2}\underline{s}_r^T H_r \underline{s}_r + \frac{1}{2}\underline{s}_f^T H_m K_J^{-1}\underline{s}_f \\ &+ \frac{1}{2}\tilde{\underline{p}}_r^T \Gamma_r^{-1}\tilde{\underline{p}}_r + \frac{1}{2}\tilde{\underline{p}}_{m1}^T \Gamma_{m1}^{-1}\tilde{\underline{p}}_{m1} \\ &+ \frac{1}{2}\tilde{\underline{p}}_{m2}^T \Gamma_{m2}^{-1}\tilde{\underline{p}}_{m2}\end{aligned} \quad (13)$$

After little algebra for the time derivative of the Lyapunov candidate function the following inequality is obtained:

$$\begin{aligned} &If\ \|\underline{s}_f\|_2 \neq 0, \dot{V}(t) \leq -\underline{s}_r^T K_r \underline{s}_r - \underline{s}_f^T K_f \underline{s}_f \\ &If\ \|\underline{s}_f\|_2 = 0, \dot{V}(t) \leq -\underline{s}_r^T K_r \underline{s}_r \end{aligned} \qquad (14)$$

From the above analysis it is concluded that $V(t)$ is a Lyapunov function and the designed system is globally asymptotically stable. It can be shown that if the matrices K_r and K_f are selected in order to satisfy the following relations:

$$\begin{aligned} &\sup(\ eig\,(K_r)) \leq \inf(\ eig\,(H_r)) \\ &\sup(\ eig\,(K_f)) \leq \inf(\ eig\,(H_m K_J^{-1})) \end{aligned} \qquad (15)$$

along the desired path, then the system is exponential stable around the desired trajectory.

3 SIMULATION RESULTS

To validate of the performance of the proposed adaptive robot controller two numerical simulation for a 2 DOF robot manipulator with flexible joints are carried out. In the first simulation the stiffness of the joints is 500 Nm/rad and the robot has to each a target point. In the second simulation the stiffness of the joints is 5 Nm/rad and the robot has to follow a desired trajectory.

The characteristics of the rigid subsystem are the following:

m_1=3Kg, m_2=3Kg, l_1=0.5m, l_2=0.5m

It is assumed that the masses m_1 and m_2 are concentrated at the distal end of each link.

The following matrices give the characteristics of the flexible subsystem for the first simulation:

$$H_m = Diag(0.3, 0.3)\ Kg$$

$$K_J = Diag(500, 500)\ \frac{Nm}{rad}$$

The chosen control parameters are given in the following matrices:

$$\Lambda_r = Diag(2, 1.7)$$

$$\Lambda_f = Diag(5.5, 5.5)$$

$$K_r = 1.2 \cdot \begin{bmatrix} 3.5 & 0.75 \\ 0.75 & 3.5 \end{bmatrix}$$

$$K_f = Diag(0.002, 0.002)$$

The matrices K_f and K_r are selected in order to satisfy the conditions (15) for an exponential stability of the system.

The adaptation gain matrices are the following:

$$\Gamma_r = Diag(0.001, 0.001, 0.001, 9, 4)$$

$$\Gamma_{m1} = \Gamma_{m2} = Diag(0.001, 0.001)$$

The above matrices are selected in order to reduce the influence of the adaptive system oscillations to the control system.

The initial values for the links positions are $(-90^o, 0^o)$ and the desired positions are $(-60^o, 30^o)$. In Fig1, Fig.2, Fig.3, and Fig.4 the position and the velocity of the links are illustrated. It is obvious that the desired positions are achieved. In Fig.5, Fig.6 are shown the joint torques reaching the steady state without undesirable oscillations.

In the second case the joint stiffness is changed and also the robot is enforced to follow a path. The stiffness matrix of the flexible subsystem is the following:

$$K_J = Diag(5, 5)\ \frac{Nm}{rad}$$

In this case the joint stiffness is the 1/100 of the first case stiffness. The used control parameters are the following:

$$\Lambda_r = Diag(2, 2)$$

$$\Lambda_f = Diag(5., 5.)$$

$$K_r = 0.26 \cdot \begin{bmatrix} 3.5 & 0.75 \\ 0.75 & 3.5 \end{bmatrix}$$

$$K_f = Diag(0.180, 0.180)$$

$$\Gamma_r = Diag(0.01, 0.01, 0.01, 1, 1)$$

$$\Gamma_{m1} = \Gamma_{m2} = Diag(0.01, 0.01)$$

In Fig.7-10 the achieved and desired positions trajectories and velocities for the two links are shown. The desired position and velocity graphs are represented with dashdot lines. Despite the very low joint stiffness the tracking error is small and continuously decreasing. The low joint stiffness and the small values of the elements of the matrix K_r are responsible for the phase shift between the achieved and the desired links angular positions. The joint torques are shown in Fig.11 and Fig.12, reaching the steady state without considerable oscillations.

In both simulations the controller achieves the desired task independently of the joints stiffness value. The chosen control values could be implemented easily because are not quite big or very small.

4 CONCLUSIONS

In this paper, a control algorithm is presented for flexible joint manipulators based on the joint torque feedback. It is shown that the physical system of equation (2) is marginally stable. Even though the control algorithm achieves global asymptotic stability for the system. The characteristics of the rigid and flexible subsystem are assumed unknown. An adaptive control algorithm based on the tracking error of the rigid and flexible subsystem estimates the values of these characteristics. From the numerical simulations, it is clear that the control algorithm can encounter manipulators with extremely high flexibility in the joints.

REFERENCES

1. Carlos Canudas de Wit, Bruno Siciliano and Georges Bastin (Eds.), "Theory of Robot Control", Springer, 1996
2. Rogelio Lozano and Bernard Brogliato, "Adaptive Control of Robot Manipulators with Flexible Joints", IEEE Transactions on Automatic Control, vol.37, no.2, pp.174-181, February 1992.
3. Ser Yong Lim, Darren M. Dawson, Jun Hu, Marcio S. de Queiroz, "An Adaptive Link Position Tracking Controller for Rigid-Link Flexible-Joint Robots Without Velocity Measurements", IEEE Transactions on Systems Man and Cybernetics, vol.27, no.3, pp. 410-425, June 1997
4. Fouad Toufic Mrad, Shaheen Ahmad, "Adaptive control of flexible joint robots using position and velocity feedback", Int. J. Control vol.55, no.5, pp.1255-1277, 1992.
5. Song, Cai, "Robust Position/Force Control", Journal of Robotic Systems, Journal of Robotics Systems, pp.430-444, 1996.
6. L. E. Pfeffer, O. Khatib and J. Hake, "Joint Torque Sensory Feedback in the Control of PUMA manipulator", IEEE Trans. RA-5, no.4, pp.418-425, 1989
7. Guoguang ZHANG, Junji FURUSHO, "Control of Robot Arms Using Joint Torque Sensors", Proceedings of the IEEE International Conference on Robotics and Automation, pp.3148-3153, April 1997.

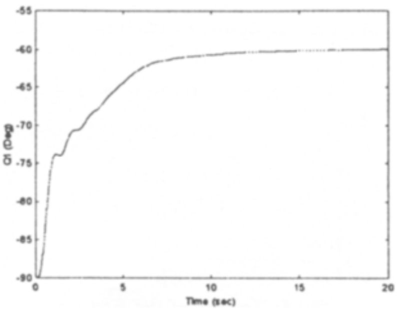


Fig.1: Position of the first link

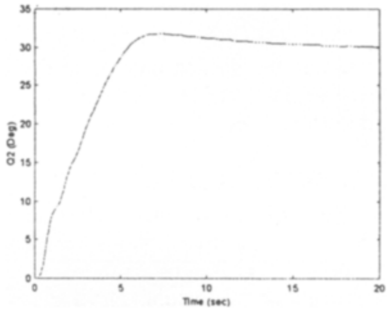


Fig.2: Position of the second link

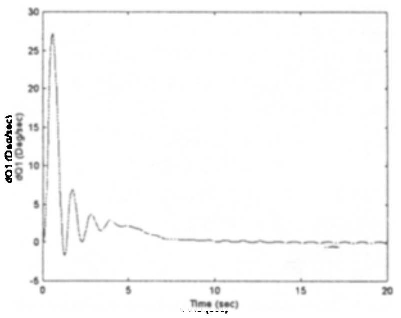


Fig.3: Velocity of the first link

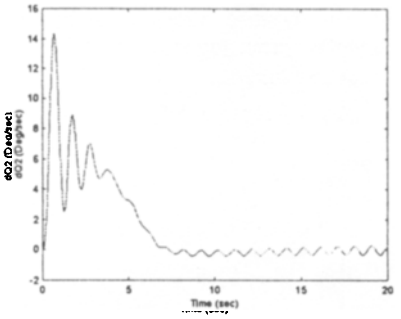


Fig.4: Velocity of the second link

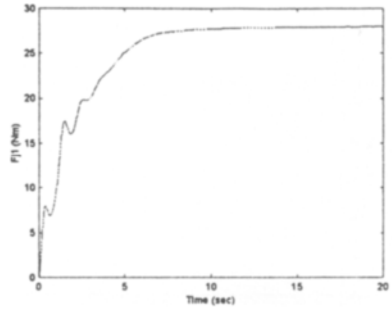


Fig.5: Torque of the first joint

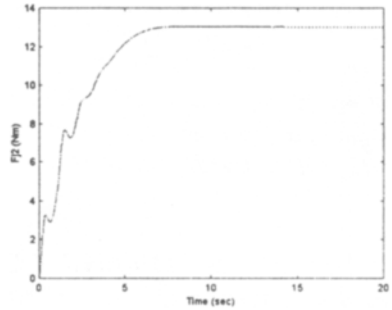


Fig.6: Torque of the second joint

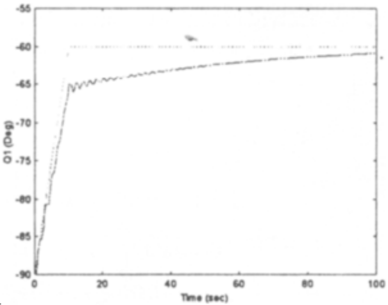


Fig7: Achieved and Desired Position of the first link

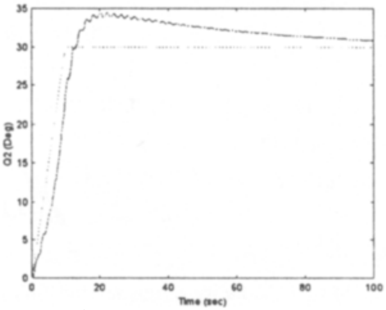


Fig.8: Achieved and Desired Position of the second link

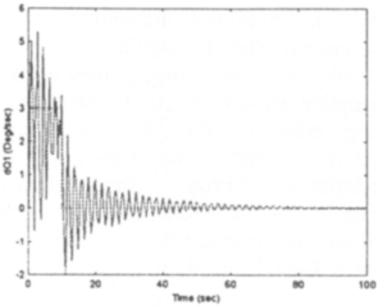


Fig.9: Achieved and Desired Velocity of the first link

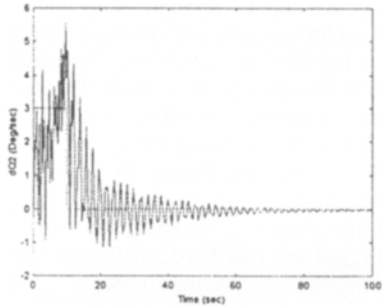


Fig.10: Achieved and Desired Velocity of the second link

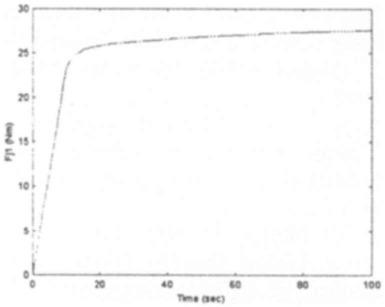


Fig11: Torque of the first joint

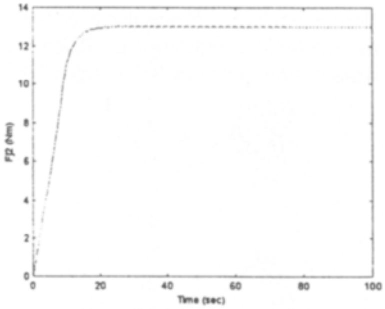


Fig.12: Torque of the second joint

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Obstacle avoidance control based on an harmonic potential field.

Svetlana A. Krasnova

Institute of Control Sciences, Profsoyuznaya 65,
117806, Mocsow, Russia, E-mail: krasnova@ipu.rssi.ru

The approach presented in this paper is based on two ideas: i) on the planning level an electrostatic potential field in configuration space of the robot is used. The generalized force curves are attractive to the goal point avoiding obstacles without local minima; ii) on the control level using the motion division method allowing decompose the problem of control syntheses on independently decided subproblems of smaller dimensionality for robot with elastic joints. Presents the control algorithms in the class of systems with discontinuous controls and high coefficients control, which provides precise tracking of the reference trajectories and the system is robust to parameter variations and external disturbances.

1.INTRODUCTION

Sliding mode control is widely used in robotics. Due to its robustness and possibilities to decouple high dimensional problem into a set of subproblems of lower dimension. In robot trajectory control with the torques as quasicontrol variables, the use of sliding mode algorithms enables efficient suppression of parameter variations and uncertainties.

In [1] was proposed a new scheme for obstacle avoidance by combining the potential method and sliding mode control. The potential method was used for planning the obstacle avoidance path and then sliding mode control made the system follow the generalized trajectory perfectly. Unfortunately, direct application of this algorithms in real life present some difficulty since the control torques's discontinuous pattern is, as a rule, non-realizable physically. To overcome this problem, in this paper the solution of problem obstacle avoidance control is proposed take into account the drives dynamics that electric drives are used as actuating elements [2].

2.THE DESCRIPTION OF DINAMIC MODEL

The motion of a rigid manipulator with n degrees of freedom and taking into account the elastic joints and the drives dynamics can be described as [3]

$$H(q)\ddot{q} + C(q,\dot{q})\dot{q} + F(q) = \tau$$
$$D\ddot{\tau} + Z\dot{\tau} + \tau = u \qquad (1)$$
$$\tau = K(a-q), D = JK^{-1}, Z = RK^{-1}$$

where $q \in R^n$ is the joint space position vector, $a \in R^n$ is the actuator position vector, $u \in R^n$ - the vector of control input to the system. It is assumed that $q, \dot{q}, a, \dot{a}$ are measurable. In the case of q, q or only q are measurable, can use the nonlinear asymptotic observers that estimates the position and speed actuator or speeds of the links [4], [5] . $H(q)$ is the $n \times n$ positive symmetrical mass or inertia matrix. $C(q,\dot{q})$ is a $n \times n$ matrix which describes centrifugal and Coriolis forces, and $F(q) \in R^n$ describes gravitational and external disturbance forces. J is the actuator inertia matrix, R is the actuator damping matrix and K is the motor - shaft stiffness matrix. While

$H(q), C(q,\dot{q}), F(q)$ are nonlinear functions, J, R, K are $n \times n$ known constant positive definite diagonal matrices. So $H(q)^{-1}, D^{-1}$, K^{-1} are existed.

3. THE DESING OF REFERENCE TRAJECTORIES BY POTENTIAL FIELD METHOD

In the system configuration space R^n the charges are placed in accordance with obstacle locations. Point charges of positive sign $C_i, i = 1,\dots,m$ is placed in the obstacle closes to the position of the point mass and a negative unit charge C^* is placed in the goal point. The idea is to use an electrostatic field as an artificial function to specify the desired trajectories. The main advantage of a Coulomb potential is the implicit absence of local minima, which would generate stable equilibrium points at undesired locations.

The potential field is a Laplace potential and it is a direct consequence of the Maxwell equation applied to static field and also known as the Gaussian property

$divE = divgradU = \frac{c}{\varsigma}$, ς is permitivity

is valid everywhere. c is the charge density and its value is zero expect for the goal and obstacle. This means that the generalized force curves are continuous and there are no local minimum and maximum away from the charges, at point where c is zero [6].

The standard Coulomb potential for point charges was extend to the n-dimensional Euclidean space $x^T = (x_1,\dots,x_n)$ by solving the Laplace equation

$$\nabla^T \nabla U(x) = \sum_1^n \frac{\partial^2 U(x)}{\partial x_i^2} = 0.$$

The generalized harmonic potential of a point charge C is given by

$$U(q) = \begin{cases} \frac{C}{q^{n-2}}, n = 1,3,4,\dots \\ C\ln\frac{1}{q}, n = 2 \end{cases} \qquad (2)$$

$$q = (\sum_{i=1}^{n} x_1^2)^{1/2}.$$

The associated electrostatic force field is

$$E(q) = -gradU(q) = \begin{cases} (n-2)\frac{C}{q^{n-1}}, n = 1,3, \\ \frac{C}{q}, n = 2 \end{cases}$$

If the obstacles is approximated by a spheres theirs centers x_i and radiuses q_i is found by geometrical consideration:

$$x_i = \arg \min_{x \in O_i} \max_{y \in O_i} \rho(x,y)$$

$$q_i = \max_{y \in O_i} \rho(x_i, y).$$

When determine the magnitudes of C_i taking into account that

i) the force curves of this field do not go to infinity

$$\sum_{i=1}^{m} C_i < C^*,$$

ii) none of the force curves of this field should be directed into ""forbidden" zone around each obstacle, that is at any point on the boundary of this zone, the gradient should be directed away from obstacle

$$\left|E(C_1, C_2, \dots, C_{i-1}, C_{i+1}, \dots, C_m, C^*)\right| + \varsigma_i <$$
$$\left|E(C_i)\right|,$$

where $\varsigma_i = \frac{1}{(Q_i - q_i)^{n-1}} - \frac{1}{Q_i^2}$.

ς_i - correction on difference electrostatic force in obstacle center and in this boundary. Q_i - minimal distance from the center of i obstacle to centers of other obstacles.

Position and velocity of the link are chosen as the system control variables so that and the system is attracted to the force field curves and it reaches the goal point avoiding obstacles.

4. Control Strategy base of the motion division method.

The motion division method in the robotics tasks [7] allows decompose the problem of control syntheses on independently decided subproblems of smaller dimensionality. In the paper this method enables the design of efficient continuous and discontinuous control algorithms, which provides precise tracking of the reference trajectories.

Let $g(t)$ be the vector of reference trajectories for the link position, which is assumed to be bounded along with its time derivatives up to order four. Defining the tracking error $e_q \in R^n$ as

$$e_q(t) = q(t) - g(t),$$
$$\dot{e}_q = \dot{q} - \dot{g}, \quad \ddot{e}_q = \ddot{q} - \ddot{g}. \qquad (3)$$

The goal of control is providence $e(t), \dot{e}(t) \to 0$.

In the case $g = U(q)$ with $U(q)$ from(2)

$$\dot{g} = v_d(t) gradU \qquad (4)$$
$$\ddot{g} = \dot{v}_d(t) gradU - v_d(t) \frac{d}{dt} gradU,$$

the system is attracted to the force field curves and reaches the goal point avoiding obstacles.

In (4) $v_d(t)$ is the scalar velocity and provide boundednees of the control

$$v_d(t) = \frac{\min(v_0, k_v r(t)^{1/2}}{\|gradU\|},$$

where v_0 is desired standard velocity, $r(t)$ is Euclidean norm of the distance to the goal point, k_v isscalar gain depending on the system deceleration capabilities. A proper choice of k_v prevents overshootat the goal point and the same guarantees a finite approach.

4.1. The stabilization with continuous control

Present the first equation from (1) by the Cauchy normal form taking into account (3):

$$\dot{e}_1 = e_2$$
$$\dot{e}_2 = H^{-1}[\tau - C[e_2 + \dot{g}] - F] - \ddot{g} \qquad (5)$$
$$e_1 = q - g, e_2 = \dot{q} - \dot{g}.$$

Next procedure makes it possible to devise the solution of system (5) stabilization into independently solvable subproplems of lesser dimension stabilization.

In the first step, assume in the first equation (5) $e_2 = -K_1 e_1$, $K_1 = diag(k_{1i})$, where k_{1i} determine the convergence rates in equation $\dot{e}_1 = -K_1 e_1$.

In the second step the problem is being independently solved to provide movement on the variety, chosen in the first step. To provide this relationship, written with respect to variables $e_2^* = e_2 + K_1 e_1$ is solved:

$$\dot{e}_2^* = H^{-1}[\tau - C[e_2^* - K_1 e_1 + \dot{g}] - F]$$
$$-\ddot{g} + K_1(e_2^* - K_1 e_1). \qquad (6)$$

System (6) stabilization problem has been solved by continuous guasicontrol

$$\tau = \tau^* = H[-K_2 e_2^* + \ddot{g} - K_1(e_2^* - K_1 e_1)] +$$
$$C[e_2^* - K_1 e_1 + \dot{g}] + F, \qquad (7)$$

where K_2 determines the convergence rates of the closed system $\dot{e}_2^* = -K_2 e_2^*$.

So, provided (6) and (7) there is the "manipulator loop" (1) stabilization.

In the third step have provide

$$e_\tau = \tau - \tau^* = 0.$$

Present the second equation of the system (1) by the Cauchy normal form concerning

$$e_3 = \tau - \tau^*, e_4 = \dot{\tau} - \dot{\tau}^*$$
$$\dot{e}_3 = e_4 \qquad (8)$$
$$\dot{e}_4 = D^{-1}[u - Z[e_4 + \dot{\tau}^*] - (e_3 + \tau^*)] - \ddot{\tau}^*$$

For stabilization (8) can use the procedure analogous (6), (7) with the continuous control:

i) $e_4 = -K_3 e_3$,

$K_3 = diag(k_{3i})$, where k_{3i} determine the convergence rates in equation $\dot{e}_3 = -K_3 e_3$.

ii) $e_4^* = e_4 + K_3 e_3$

$$\dot{e}_4^* = D^{-1}[u - Z[e_4^* - K_3 e_3 + \dot{\tau}^*] -$$

$$(e_3 + \tau^*)] - \ddot{\tau}^* + K_3(e_4^* - K_3 e_3).$$

iii) $$u = D[-K_4 e_4^* + \ddot{\tau}^* - K_3(e_4^* - K_3 e_3)] + Z[e_4^* - K_3 e_3 + \dot{\tau}^*] + (e_3 + \tau^*), \qquad (9)$$

where K_4 determines the convergence rates of the closed system $\dot{e}_4^* = -K_4 e_4^*$.

Thus, the proposed procedure by means of variables nonsingular change and control choice (9) allows to solve the stabilization task in (5) and (8) with linear feedback and to appoint desirable convergence rates the state variables independently in every subsystem by choice K_1, K_2, K_3, K_4.

Note that realization of compensating algorithm (9) may happen to be too complicated because of the necessity to obtain sufficiently accurate estimates of object parameters. Unfortunately , the exact values of them, as rule, are not available. To provided robust control performance despite unknown disturbances and parameter variations, having the method of the theory of systems with discontinuous control [8] . On the other hand, in the case of electric drives controlling action u (for example, its are voltage) have discontinuous due to using semiconductor switch devices [9]. So for stabilization (5), (8) preferab to use the discontinuous algorithms of control as the most simple from the viewpoint realization.

4.2. The stabilization with discontinuous control

The presented the control algorithms in the class of systems with discontinuous and high gain control allow decompose the problem of control syntheses on independently decided subproblems of smaller dimensionality. In this case the system is robust to parameter variations and external disturbances.

It is assumed that elements of object parameters matrixes are measuredd with the known error: $H^{-1} = H_0 + \Delta H$, $C = C_0 + \Delta C$, $F = F_0 + \Delta F$, where H_0, C_0, F_0 are accessible for measurements. $\Delta H, \Delta C, \Delta F$ are unknown, constrained by module functions $|\Delta H| \le \overline{H}$, $|\Delta C| \le \overline{C}$, $|\Delta F| \le \overline{F}$, $\overline{H}, \overline{C}, \overline{F}$ - const.

So the system (5) is follows as

$$\dot{e}_1 = e_2 \qquad (10)$$

$$\dot{e}_2 = H_0[\tau - C_0[e_2 + \dot{g}] - F_0] - \ddot{g} + Y,$$

where

$$Y = \Delta H[\tau - C[e_2 + \dot{g}] - F] -$$

$$H_0 \Delta C[e_2 + \dot{g}] - H_0 \Delta F$$

$$|Y| \le \overline{Y}, \overline{Y} \text{ -const.} \qquad (11)$$

Sliding mode control reduce the system dimensionality and enables the design of efficient control algorithms for obstacle avoidance by making the system trajectories follow the force curves precisely. If we use the following sliding manifold for system (5)

$$s_q(q_1, q_2) = e_2 + l_q e_1 = 0$$

then after sliding mode starts the system trajectory is assured to be one of the force curves and it reaches the goal point avoiding obstacles.

In the first step to make sliding mode occur on the manifold $s_q = 0$, the standard approach can be use. If to choose the quasicontrol as

$$\tau = \tau^* = -U_q sign(s_q),$$

$$sign(s_q)^T = (sign(s_{q1}), \ldots, sign(s_{qn})) \qquad (12)$$

where U_q is a diagonal matrix with positive elements, then the condition for sliding mode existence $s_{qi}\dot{s}_{qi} < 0$ may be fulfilled. But the quasicontrol force defined in (12) is discontinuous at $s_q = 0$ and thus not feasible with real - life actuators due to their inertia. Therefore it is proposing to use a continuous approximation [10] $f(s_q)$ of discontinuous control $sign(s_q)$ by piece - wise linear approximation

$$f(s_q) = sat(s_q) = \begin{cases} 1, s_q > \varepsilon \\ s_q / \varepsilon, \left|s_q\right| < \varepsilon, \varepsilon > 0 \\ -1, s_q < -\varepsilon \end{cases}, (13)$$

When

$\varepsilon \to 0, f(s_q) \to sign(s_q)$ corresponds to the ideal sliding mode.

Let the quasicontrol is follows as

$$\tau = \tau^* = H_0^{-1}(-U_q f(s_q) - U_{qq}), \qquad (14)$$

$$f^T(s_q) = (f(s_{q1}), \dots, f(s_{qn})),$$

$$U_{qq}^T = (U_{qq1}, \dots, U_{qqn})$$

where

$$U_{qq} = H_0[-C_0[e_2 + v_d(t) gradU] - F_0] -$$

$$\dot{v}_d(t) gradU - v_d(t)\frac{d}{dt} gradU + l_q e_2 (15)$$

compensate the available for mesuring part in (10).

The derivative of s_q taken into account (4), (10), (14) is

$$\dot{s}_q = \dot{e}_2 + l_q \dot{e}_1 = -U_q f(s_q) + Y \qquad (16)$$

To ensure the existence of sliding mode in the manifold $s_q = 0$, we examine the Lyapunov function candidate

$$V_q = \frac{1}{2} s_q{}^T s_q, \qquad (17)$$

The derivative of V_q for the ideal sliding mode $\varepsilon \to 0$ is expressed as

$$\dot{V}_q = s_q^T \dot{s}_q = s_q^T(-U_q sign(s_q) + Y) \ . \qquad (18)$$

For $\dot{V}_q$ an upper estimate can be found to satisfy the following inequality

$$\dot{V}_q \le \left|s_q\right|(-U_q + \overline{Y}). \qquad (19)$$

That is, if the control gain vector U_{qi} is selected to be $U_q > \overline{Y}$, then $\dot{V}_q < 0$ and sliding mode occurs. l_q determine the convergence rates in equation $\dot{e}_1 = -l_q e_1$.

In up limit case, when $\varepsilon > 0$ the system stabilization is possible only with definite accuracy $\left|s_q\right| \le \Delta, \Delta = O(\varepsilon)$. .

The derivative of V_q is expressed as

$$\dot{V}_q = s_q^T(-U_q f(s_q) + Y). \qquad (20)$$

With (13) an upper estimate for $\dot{V}_q$ from (20) is

$$\dot{V}_q \le \left|s_q\right|^T(-U_q\left|s_q\right| / \varepsilon + \overline{Y}).$$

That is, then $\dot{V}_q < 0$ outside the district $\left|s_q\right| > \Delta = \frac{\overline{Y}\varepsilon}{U_q}$ and consequently the state vector converges in the district

$$\left|s_q\right| \le \Delta = \frac{\overline{Y}\varepsilon}{U_q}, \left|e_1\right| \le \frac{\Delta}{l_q}. \qquad (21)$$

For any electric drives the torques τ is bounded and the choice U_q depends on the control syntheses in "actuator loop" in the second step.

Note, that the control syntheses in "actuator loop" demands $\dot{\tau}^*, \ddot{\tau}^*$ to be bounded. When $f(s_q) = sat(s_q)$ (13), the derivative of $f(s_q)$ is not bounded in singular points $\left|s_q\right| = \varepsilon$. To overcome this problem, we may use the unit control approximation of $sign(s_q)$

$$f(s_q) = \frac{s_q}{\left\|s_q\right\| + \varepsilon}, \varepsilon > 0$$

In this case the derivative of $f(s_q)$ is bounded everywhere.

In the second step fulfillment of relation $\tau - \tau^* = 0$ with τ^* from (14) can be achieved within a finite interval of time due to syntheses of the sliding movements on the variety $s_\tau = e_4 + l_\tau e_3 = 0$ with discontinuous control

$$u = -DU_\tau sign(s_\tau), \tag{22}$$

$$sign(s_\tau)^T = (sign(s_{\tau 1}),\dots,sign(s_{\tau n}))$$

where U_τ is a diagonal matrix with positive elements and it depends on supple voltage of electric drives.

The equivalent control from $\dot{s}_\tau = \dot{e}_4 + l_\tau \dot{e}_3 = 0$ with (8) is

$$u_{eq} = Z[e_4 + \dot{\tau}^*] + (e_3 + \tau^*) + D\ddot{\tau}^* - l_\tau e_4$$

If $\dot{\tau}^*, \ddot{\tau}^*$ are bounded, an upper estimate ω_τ can be found to satisfy the following inequality

$$\left|D^{-1}u_{eq}\right| \le \omega_\tau, \omega_\tau{}^T = (\omega_{\tau 1},\dots,\omega_{\tau_n}) \tag{23}$$

Taking into account (22) $\dot{s}_\tau$ follows as

$$\dot{s}_\tau = -U_\tau sign(s_\tau) - D^{-1}u_{eq} \tag{24}$$

To ensure the existence of sliding mode in the manifold $s_\tau = 0$, we examine the Lyapunov function candidate

$$V_\tau = \frac{1}{2} s_\tau{}^T s_\tau, \tag{25}$$

$$\dot{V}_\tau = s_\tau{}^T \dot{s}_\tau = s_\tau^T(-U_\tau sign(s_\tau) - D^{-1}u_{eq}).$$

With (23) an upper estimate for $\dot{V}_\tau$ is

$$\dot{V}_\tau \le \left|s_\tau\right|^T(-U_\tau + \omega_\tau). \tag{26}$$

that is, if $U_\tau > \omega_\tau$ (27) then $\dot{V}_\tau < 0$ and sliding mode occurs. l_τ determines the convergence rates in equation $\dot{e}_3 = -l_\tau e_3$.

Note that, unlike continuos systems, the switching in control excite the unmodeled dynamics, which leads to chattering (the oscillations in the state vector at a finite frequency). A recent study [2] and practical experience show that the chattering caused by unmodeled dynamics may be eliminated in systems with asymptotic observers.

From (23) and (27) follows that is the maximum value for U_q defined as nonlinear function of U_τ

$$U_q \le U_{q\max}(U_\tau). \tag{28}$$

Substitute (28) into (21) the minimal possible convergence district of state vector (5)

$$\left|e_1\right| \le \frac{\Delta_{\min}}{l_q}, \Delta_{\min} = \frac{\overline{Y}\varepsilon}{U_{q\max}}$$

is obtained.

Thus, the proposed procedure allows to provide the robust control performance with maximum feedback coefficients taking into account the boundednees of drives torques.

5. CONCLUSION

This study presents an global dynamic output feedback controller for robots with flexible joints taking into account the drives dynamics, based on the motion division method. The robustness properties and possibilities to reduce system order to a lower order are obtained by sliding mode, which makes the state trajectories identical to the force curves. The generalized force curves are attractive to the goal point avoiding obstacles without local minima.The stabilization accuracy of the closed - loop system is examine by Lyapunov second method.

REFERENCES

1.H.Hashimoto, F. Harashima, I. Kalico, S.Krasnova, V. Utkin. Sliding mode control and Potential field in obstacle avoidance. Proceedings of ECC'93 held in Groningen, Holland, June 28 - July 1, 1993, vol.3, pp.1544-1548,1993.

2. V.I.Utkin. Sliding mode control design principles and applications to electric drives. IEEE Transactions industrial electronics, vol.40, no.1, pp. 23-36, February 1993.

3. J.Yuan and Y.Stepanenko. Composite adaptive control of flexible joint robots.

Automatica, vol. 29, no.3,pp.609-619,1993.Printed in Great Britain.

4. S. Nicosia and P. Tomei. A global output feedback controller for flexible joint robots. Automatica. vol. 31, no. 10, pp. 1465 - 1469, 1995.

5. A. Lukyanov A.G. and Dodds S.J. Non-linear state observer design for block control with sliding mode control of manipulator. Proceeding of the EFFACE Workshop "Motion control".1995. Munich, Germany. pp.632 - 639.

6. H. Noborio, S. Wazumi, S. Arimoto. implicit approach for robot raining on a force field without generation of local minima. Preprints of the 11-th IFAC Congres, Tallin, USSR, pp.85-90, 1990.

7. V.A. Utkin. Constrained robot control based on the method of movements separation. Tampere International Conference on Machine Automation Mechatronics Spell Profitabiliti, Tampere,pp.173-180, Finland,February, 1994.

8. V.I.Utkin. Sliding modes in optimization and control. New York, Springer-Verlag,1992.

9. V.A. Utkin. The control problems of induction motor. Automatics and Telemechanics, no.12,1993.

10. J.E. Slotine. Sliding controller design for non-linear systems. Int. J. Control, vol. 40, no.2, pp.421-434, 1984.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Smooth Robust Control for Robot Manipulators

B W Bekit, J F Whidborne and L D Seneviratne
Department of Mechanical Engineering, King's College London, Strand, London WC2R 2LS,UK.

This paper proposes an algorithm for smoothing the control output of a sliding mode controller which has the capability to reduce or remove undesirable chattering while keeping the robust characteristic that reject system uncertainties. Simulation results show that the proposed controller gives a good system performance in the face of uncertain system parameters and external disturbances.

1. Introduction

A controller designed for robotic systems must be able to cope with all kinds of changes, whether they are highly dynamic, slowly varying, or persistent, whether they are known in advance, measured or unknown. From the point of view of control system design the system should be robust, or insensitive, with respect to such changes, which can be interpreted as disturbances to or uncertainty about the behaviour of the system.

Robotic manipulators are non-linear, highly coupled multivariable systems which pose many practical problems in terms of control. Today, conventional fixed-gain controllers are extensively used in industrial robots. However, this controller becomes less effective when high speed is required with high position accuracy. Therefore, there is a need for a sophisticated controllers, which could automatically adapt itself to changes in the robot dynamics and perform satisfactorily in the presence of uncertainties.

Adaptive control has the potential to give desired performance for system with uncertainty. The controller has the ability to adjust itself from knowledge, successively gained about the robot manipulator. This controller perform well in the presence of structured uncertainty. However, it is sensitive to input and output noise, and unmodeled dynamics.

Robust control guarantees stability and performance for uncertain systems. The advantages of robust control over adaptive control are computational simplicity in implementation, better compensation for time-varying parameters and for unstructured nonlinear uncertainties, and guaranteed stability. The variable structure design approach, which includes the sliding mode method [6] is one of the popular robust control approaches. The main drawback of the variable structure control methods is that of chattering where the controller input undergoes very high speed switching. The extent of the chattering problem is dependent on the controller gains. In general, the larger the gains the more serious the chattering problem.

Many researchers have addressed the problem of chattering. For example In [5], the discontinuous control is approximated inside a boundary layer located around the switching surface.

In this paper, a smooth robust controller for a robot manipulator is developed. The switching gains are incorporated in the sliding mode controller to accelerate the state trajectories toward the sliding hyperplane.

In order to evaluate the proposed robot control scheme, extensive computer simulations are performed on a two degree-of-freedom planar robot manipulator. Simulation results show that the proposed robot control scheme is able to successfully react to sudden changes in payload, dynamic uncertainties and time-varying manipulator parameters.

2. Robot manipulator dynamics

Using a Lagrangian formulation, the generalised forces for an n-link manipulator excluding the friction and other disturbance can be expressed as a second-order non-linear differential equation,

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \tau, \tag{1}$$

where q, $\dot{q}$ and $\ddot{q}$ are $\mathbb{R}^n$ vectors of the joint position, velocity and acceleration respectively. $M(q) \in \mathbb{R}^{n\times n}$ is a pseudo-inertia symmetric matrix, $C(q) \in \mathbb{R}^n$ is a vector of non-linear coriolis and centrifugal terms, $G(q) \in \mathbb{R}^n$ is vector of gravity effect, and $\tau \in \mathbb{R}^n$ is the vector of joint torque.

For the subsequent discussion, the following well-known properties of dynamics of robot manipulator are introduced.

Property 1 *The matrix* $\dot{M}(q) - 2C(q,\dot{q})$ *is skew symmetric,* $x^T[\dot{M}(q) - 2C(q,\dot{q})]x = 0 \quad \forall x \in \mathbb{R}^n$.

Property 2 *The structure of dynamic equation (1) is linear in terms of a suitable selected vector of robot and load parameters,*

$$\tau = M(q)\ddot{q} + C(q,\dot{q})\dot{q} + G(q) = \psi(q,\dot{q},\dot{q},\ddot{q})\theta,$$

where $\psi(q,\dot{q},\dot{q},\ddot{q}) \in \mathbb{R}^{n\times r}$ *is the known non-linear function of the manipulator structural dynamics,* $\theta \in \mathbb{R}^r$ *is the vector containing the unknown manipulator parameters. These include the masses and moment of inertia of the links, the payloads and the link dimension.*

3. Robust control

The simplest procedure to design a robust controller is to extract a known, nominal controller from the perfect control. What is left in the perfect control law after this extraction are the terms that are unknown due to the lack of perfect knowledge of the system. Furthermore, since M, C and G can be made linear in terms of the manipulator parameters, Property 2, reordering (1) gives an equation where θ denotes the vector of p unknown parameters, and ψ_u, ψ_k denotes functions of signal variables [2];

$$\tau = \psi_u(q,\dot{q},\dot{q}_r,\ddot{q}_r)\theta + \psi_k(q,\dot{q},\dot{q}_r,\ddot{q}_r), \tag{2}$$

where $\psi_u \in R^{n\times p}, \psi_k \in \mathbb{R}^n$ are vectors of signal variables, $\tau \in \mathbb{R}^n$ is the vector of joint torque and $\theta \in \mathbb{R}^p$ denotes the vector of p unknown parameters. The reference velocity vector, $\dot{q}_r$ is defined as follows

$$\dot{q}_r(t) = \dot{q}_d(t) - \lambda\tilde{q}(t) \tag{3}$$

where λ is strictly positive constant. Hence, the vector s can be written as,

$$s(t) = \dot{q}(t) - \dot{q}_r(t). \tag{4}$$

Equation (4) represents a notation manipulation that allows one to translate energy- related properties into trajectory control properties. When the trajectory errors are confined to the sliding surface, the system exhibits desirable behaviour. The resulting system equations are given by n decoupled first order linear differential equations,

$$\dot{\tilde{q}}_i(t) = -\lambda_i\tilde{q}_i(t) \quad i = 1,2,\ldots,n. \tag{5}$$

In robot manipulator control each equation represents the dynamics of a single degree of freedom when the robot is in sliding mode. Thus, by controlling the robotic system to move in sliding mode the non-linear interaction in its dynamics are completely eliminated. The dynamic performance in sliding mode depends on the design parameters λ_i; choosing the λ_i terms large and positive speeds the motion in sliding mode.

If the parameter θ is known, we can use the following control law to achieve perfect tracking [4],

$$\tau(t) = \psi_u(q,\dot{q})\theta + \psi_k(q,\dot{q}) - K_v s, \tag{6}$$

where $K_v = K_v^T > 0$. However, the assumption that all parameters of the system are known is generally not valid. Not only are the parameters not known or poorly known, but they may also be subject to change as the manipulator performs certain tasks. If a manipulator picks up an unknown load we may treat the load as part of another link. Since the parameters of (6) depend on the manipulator structure and payload it carries, it is difficult to obtain completely accurate values for these parameters. We can get estimated value of manipulator payload parameters. This suggests that equation (6) can be written as:

$$\tau(t) = \psi_u(q,\dot{q},\dot{q}_r,\ddot{q}_r)\tilde{\theta} + \psi_k(q,\dot{q},\dot{q}_r,\ddot{q}_r) - K_v s \tag{7}$$

where $\tilde{\theta} = \hat{\theta} - \theta$ is the parameter estimation error. We are interested in deriving a sliding mode control law τ such that for any uncertainty in the parameters of the robotic system, $\tilde{q}(t) \to 0$ as $t \to \infty$. The control objective is to make the robot track a given desired trajectory $q(t) = q_d(t)$. In this case, a simple and common choice of sliding mode surface is that of $s = 0$, for some $\lambda > 0$.

From the robot manipulator equation (1) and using (4) and Property 2, we get

$$\begin{aligned} &M(q)\dot{s} + C(q,\dot{q})s \\ &+\psi_u(q,\dot{q},\dot{q}_r,\ddot{q}_r)\theta + \psi_k(q,\dot{q},\dot{q}_r,\ddot{q}_r) = \tau \end{aligned} \qquad (8)$$

Let us consider the dynamic equation (1) with control law (7) and with estimated parameters $\tilde{\theta}$ replaced with v_i , then

$$\tau(t) = \psi_u(q,\dot{q},\dot{q}_r,\ddot{q}_r)v_i + \psi_k(q,\dot{q},\dot{q}_r,\ddot{q}_r) - K_v s, (9)$$

where v_i are switching functions designed according to the variable structure theory.

Theorem 1 *If the control law (9) is applied to the robot manipulator described by equation (1) with the following continuous control law*

$$v_i = -\bar{\theta}_i \operatorname{sgn}\left(\sum_{j=1}^{n} s_j \phi_{ji}\right), \quad i = 1,2,\ldots,m \quad (10)$$

where $-\bar{\theta}_i \leq |\theta_i|$ and θ are the upper bounds of the unknown parameters, the tracking errors, $\tilde{q}(t)$ and $\dot{\tilde{q}}(t)$ exponentially converge to zero.

Proof: Let us consider the following positive definite function

$$V(t) = \frac{1}{2}[s^{\mathrm{T}} M s]. \qquad (11)$$

Differentiating (11) with respect to time and using Property 1 and using the relationship between s, q and q_r and using (8), we have:

$$\dot{V}(t) = s^T[\tau - Cs - \psi_u\theta - \psi_k] + s^{\mathrm{T}} C s.$$

Substituting control law (9), we have

$$\dot{V}(t) = s^T[\psi_u v - \psi_u\theta - K_v s]$$

where K is a positive definite matrix. Thus

$$\begin{aligned} \dot{V}(t) &= -s^T K_v s - \sum_{i=1}^{m} \bar{\theta}_i \left|\sum_{j=1}^{n} s_j \phi_{ji}\right| \\ &\quad - \sum_{i=1}^{m} \theta_i \sum_{j=1}^{n} s_j \phi_{ji}, \\ &\leq -s^T K_v s, \end{aligned}$$

and taking vector norms gives $\dot{V}(t) \leq -\lambda_{\min}(K_v)||s||^2$.

Thus $\tilde{q}(t)$ and $\dot{\tilde{q}}(t)$ exponentially converge to zero. From equations (11) and (12), since $\dot{V}(t) \leq 0$ and $V(t) \geq 0$, $V(t)$ tends to a constant as $t \to \infty$ and it is therefore upper bounded. Given (11) for V, since M is uniformly positive definite, therefore s is bounded, which implies that q and $\dot{q}$ are bounded. This implies that $s \to 0$ as $t \to \infty$, and thus $\tilde{q} \to 0$ as $t \to \infty$. Therefore the exponential convergence of the tracking error are guaranteed. □ □

3.1. Smoothing control law

The control scheme given above is discontinuous and it is well known that synthesis of such a control law gives rise to chattering. Chattering is undesirable in most control applications such as robot control since it involves high frequency control activity and may excite neglected high frequency dynamics or lead to increased wear-and tear for actuators. One of the causes of chattering is that the constant gain K may be unfavourably tuned. The larger the gains the more serious the chattering problem. To overcome control chattering, smoothing techniques have to be employed. This is achieved by smoothing the control discontinuity in a thin boundary layer neighbouring the switching surface as suggested by [4].

Let us replace the signum function in (10) with the $\operatorname{sat}(\frac{s}{\beta})$ function as suggested by [5]

$$\operatorname{sat}\left(\frac{s}{\beta}\right) = \begin{cases} -1, & \text{if } \frac{s}{\beta} < -1; \\ \frac{s}{\beta}, & \text{if } \frac{s}{\beta} \in [-1,+1]; \\ 1, & \text{if } \frac{s}{\beta} > 1. \end{cases} \qquad (12)$$

With this algorithm the sliding mode control law given (10) can be written as

$$v_i = -\bar{\theta}_i \operatorname{sat}\left(\sum_{j=1}^{n} s_j \phi_{ji}\right), \quad i = 1,2,\ldots,m. \quad (13)$$

The control law (10) causes the trajectory of the dynamic system to converge to a boundary-layer set

$$\mathrm{B}(\mathrm{s}(\mathrm{x}),\beta) \triangleq \{\mathrm{x} \in \mathbb{R}^{\mathrm{n}} : ||\mathrm{s}(\mathrm{x})|| \leq \beta\}$$

in finite time. The smoothing is achieved at the expense of tracking precision. Thus, the larger the boundary layers the smaller the control chatter and the greater the tracking error. If $|s| > \beta$, the trajectories will be directed toward the switching planes. However, if $|s| < \beta$, there is no guarantee that the trajectories will approach the sliding surface. Moreover, a fixed gain often gives too much energy for the purpose of trajectory tracking. Hence a modification of switching control is presented.

3.2. Modified smoothing control algorithm

The smoothing algorithm presented above has some drawbacks. Although this method can attenuate the degree of high-frequency control input, its stability is guaranteed only outside the boundary layer, and its asymptotic tracking often cannot be achieved if the boundary layer is insufficiently small. Moreover, a fixed switching gain often consumes too much energy for the purpose of trajectory tracking. To improve this, a new smoothing algorithm is presented which posses powerful smoothing capability to reduce or remove undesirable chattering while keeping the robust characteristic that reject system uncertainties and external disturbance.

Theorem 2 *Consider the dynamic equation (1) and the control law (7) with the parameter estimation modified to*

$$v_i = \bar{\theta}_i \frac{\delta \sum_{j=1}^{n} s_j \phi_{ji}}{|\sum_{j=1}^{n} s_j \phi_{ji}| + \sigma e^{-\beta t}}$$

Then, all the signals in the system remain bounded and tracking errors $\tilde{q}$ converge to zero exponentially.

Proof: Using the same Lyapunov function candidate equation (11) yields

$$\dot{V} = -s^T K_v s - s^{\mathrm{T}} \psi_u [v_i - \bar{\theta}].$$

Thus, if $|s_j \phi_{ji}| \leq 0$ we get

$$\begin{aligned} \dot{V} &= -s^T K_v s \\ &\leq -\lambda_{min}(K_v) ||s||^2 \end{aligned}$$

while, if $|s_j \phi_{ji}| > 0$ we get

$$\dot{V} = -s^T K_v s - \sum_{j=1}^{n} \left[\delta \frac{(s_j \phi_{ji})^2}{|s_j \phi_{ji}| + \sigma e^{-\beta t}} \bar{\theta}_i \right] - |s_j \phi_{ji}| \theta]$$

where $\sigma, \beta > 0$. Hence

$$\dot{V} \leq 2\lambda_{min}(K_v) ||s||^2 - |\bar{\theta}| \sigma e^{-\beta t}.$$

Comparing the argument of [3] we can show that $\dot{V} < 0$ for $||s|| > \left[\frac{w}{2\lambda_{min}(K_v)} \right]$, where $w = \sigma |\theta|$. □ □

3.3. Effect of bounded input disturbance

In this section, we will examine the robustness properties of the controller. Consider the dynamic equation (1) with the addition of disturbance τ_d

$$M(q)\ddot{q} + C(q)\dot{q} + G(q) + \tau_d = \tau$$

where $\tau_d \in \mathbb{R}^n$ is the vector of uncertainties presenting friction, torque disturbance etc. Substituting into equation (7) gives

$$\begin{aligned} & M(q)\ddot{q} + C(q)\dot{q} + G(q) + \tau_d \\ = \; & \psi_u(q, \dot{q}, \dot{q}_r, \ddot{q}_r) v + \psi_k(q, \dot{q}, \dot{q}_r, \ddot{q}_r) - K_v s. \end{aligned} \tag{14}$$

Let us prove the boundedness of τ_d and thereby derive the estimation law for τ_d. Consider the Lyapunov function

$$V = \frac{1}{2} s^T M s + \frac{1}{2} \tilde{\tau}_d{}^T K_I^{-1} \tilde{\tau}_d.$$

Differentiating V with respect to time and using Property 1 gives

$$\dot{V} = s^T M \dot{s} + \frac{1}{2} s^T \dot{M} s + \tilde{\tau}_d{}^T K_I^{-1} \dot{\tilde{\tau}}_d.$$

Evaluating $\dot{V}$ along the trajectory of (15) yields

$$\begin{aligned} \dot{V} &= s^T (\psi_u(q, \dot{q}, \dot{q}_r, \ddot{q}_r) v \\ &\quad + \psi_k(q, \dot{q}, \dot{q}_r, \ddot{q}_r) - K_v s - \tau_d) + \tilde{\tau}_d{}^T K_I^{-1} \dot{\tilde{\tau}}_d, \\ \dot{V} &\leq -s^T K_v s + s^T \tau_d + \tilde{\tau}_d{}^T K_I^{-1} \dot{\tilde{\tau}}_d, \\ \dot{V} &\leq -s^T K_v s - \tau_d (s - K_I^{-1} \dot{\hat{\tau}}_d). \end{aligned}$$

As d is constant we can assume $\dot{\hat{\tau}}_d = \dot{\tilde{\tau}}_d$ is this suggests defining $\dot{\hat{\tau}}_d$ as

$$\dot{\hat{\tau}}_d = K_I s(t).$$

It then follows that

$$\dot{V} \leq -s^T K_v s.$$

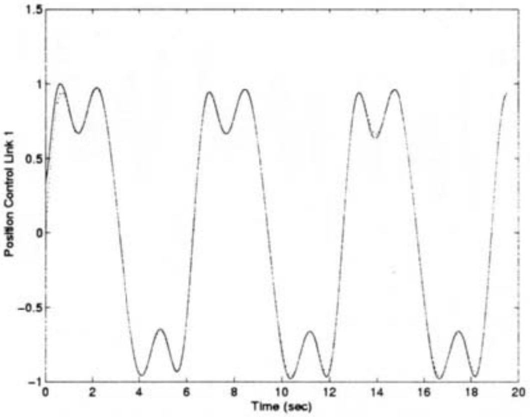


Figure 1. Position response of modified algorithm

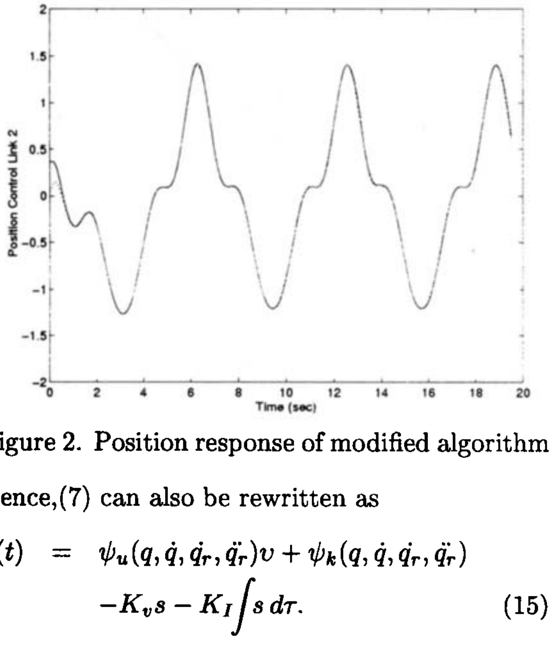


Figure 2. Position response of modified algorithm

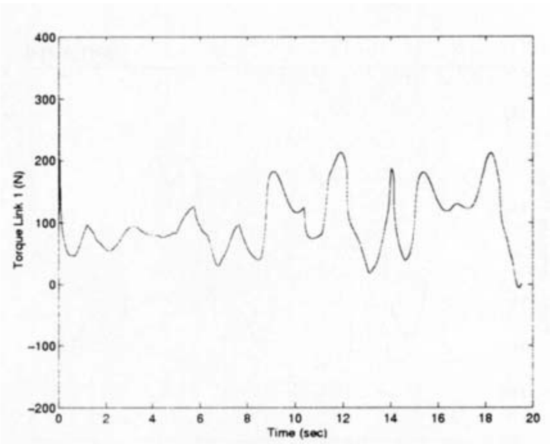


Figure 3. Torque of modified algorithm

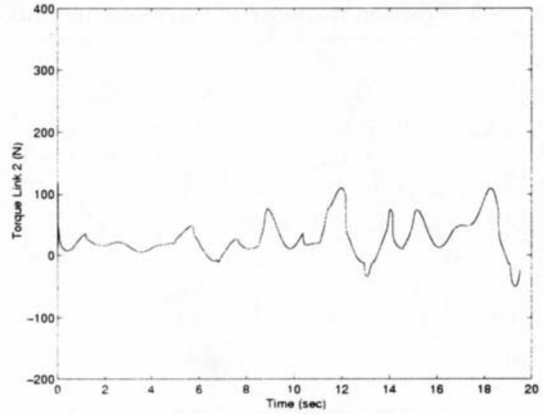


Figure 4. Torque of modified algorithm

Hence,(7) can also be rewritten as

$$\begin{aligned}\tau(t) &= \psi_u(q,\dot{q},\dot{q}_r,\ddot{q}_r)v + \psi_k(q,\dot{q},\dot{q}_r,\ddot{q}_r) \\ &\quad -K_v s - K_I\int s\,d\tau. \qquad (15)\end{aligned}$$

4. Computer simulation results

The complete set of non-linear, coupled differential equations (1) was used to model the manipulator dynamics. In this simulation it was assumed that the system dynamics are known except for an unknown payload. The objective of the control design is to make the joint position track the following desired trajectories which is shown in Figure [1-2] by the doted lines:

$$\begin{aligned}q_{1d} &= \sin(t) + 0.3\sin(3t) + 0.1\sin(4t) \\ q_{2d} &= -1.4e^{-t} + \cos(t) + 0.3\cos(3t) + 0.1\cos(4t)\end{aligned}$$

using the control law (15) and with manipulator parameter of [1]. The simulation was run with a slight initial perturbation of the joint position and velocities of $q_1(0) = -1.5706$, $\dot{q}_1(0) = 0.0$, $q_2(0) = 1.10^{-4}$, $\dot{q}_2(0) = 0.0$. We assume the load as part of link two, in which case only the parameters of the mass of the second link (m_2), length of the second link (l_{c2}) and moment of inertia of the second link (I_2) would be unknown. Then, the parameters m_2, l_{c2} and I_2 change to $m_2 + \Delta m_2$, $l_{c2} + \Delta l_{c2}$ and $I_2 + \Delta I_2$. This was tested by changing the manipulator mass of link two by Δm_2 =8.64kg, inertia by ΔI_2=0.323kgm^2 and Δl_{c2}=1.84m at $t=$ 8. The gain matrix are selected as follows $K_v = 10I_{2\times 2}, K_I = I_{2\times 2}, \sigma = I_{3\times 3}, \beta = I_{3\times 3}$ and $\delta = \{1,1,3\}$.

Figures [1-4] shows a response of the robot using the modified sliding mode control law, these clearly demonstrate the significant effect of picking unknown payload and the response of the slid-

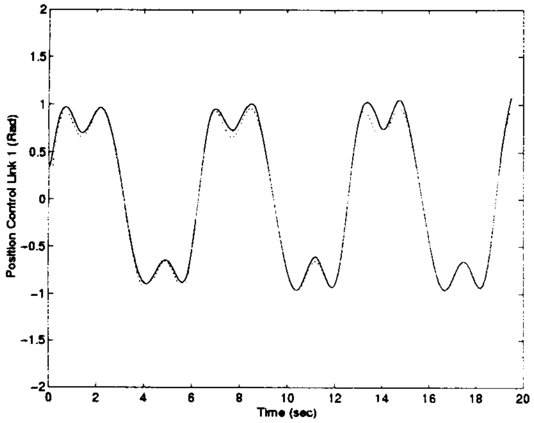


Figure 5. Position response of conventional SMC

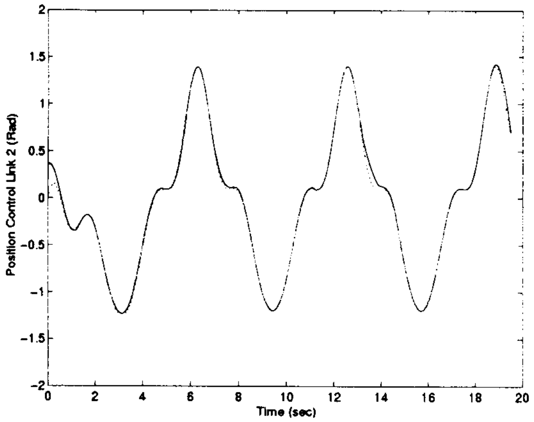


Figure 6. Position response of conventional SMC

ing mode controller. One can also see that the chattering phenomenon, which always occurs in conventional sliding mode controller design, has been effectively reduced (Figures [7-8]) and with much better tracking accuracy.

5. Conclusion

In this paper, a robust sliding mode control scheme for rigid robotic manipulator has been proposed. This technique gives a satisfactory system performance in the face of uncertain systems parameters and external disturbances. Simulation results show that the proposed scheme remove the undesirable chattering while achieving a better tracking accuracy.

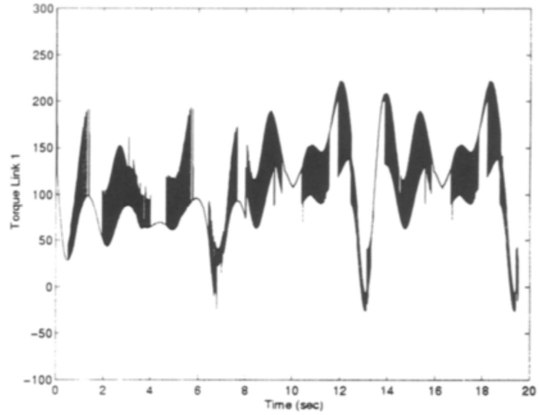


Figure 7. Torque of of conventional SMC

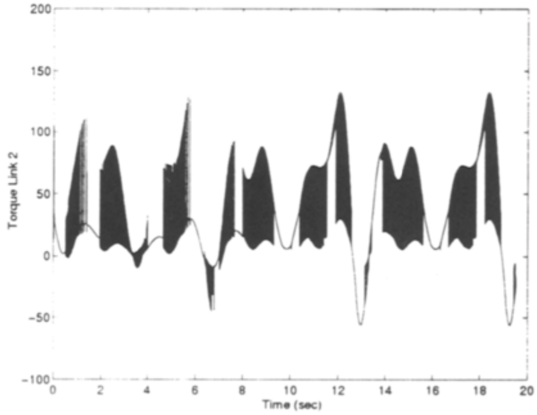


Figure 8. Torque of of conventional SMC

REFERENCES

1. Bekit B.W., Whidborne J.F., and Seneviratne L.D. Adaptive control for a robot manipulator based on hyperstabilty design. *IEEE/ASME Int. Conf. on Advance Intelligent Mechatronics, Tokyo*, June 1997.
2. R. Johansson. Adaptive control of robot manipulator motion. *IEEE Trans. Robotics and Autom.*, 6(3):483–490, 1990.
3. G. Leitman. On the efficacy of nonlinear control in uncertain linear systems,. *ASME J. Dyn. Syst. Meas. & Control*, 102:95–102, 1981.
4. J. J. E. Slotine. Sliding controller design for non-linear system. *Int. J. Control*, 40:421–434, 1984.
5. J. J. E. Slotine and W. Li. *Aplied Nonlinear Control.* Prentice-Hall, Englewood Cliffs, N.J., 1991.
6. V. I. Utkin. Variable structure systems with sliding modes–a survey,. *IEEE Trans. Autom. Contr.*, 22(3):212–222, 1977.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Robot Arm Positional Deflection Control with a Laser Light Sensor

Michael G. Puopolo, M.Sc. and Saeed B. Niku, Ph.D.
Mechanical Engineering Department
California Polytechnic State University
San Luis Obispo, California 93407

A control device was designed and constructed to actively control the tip position of a semi-flexible robotic arm and to correct the deflections of the arm in real time. The system corrects the position of the robot arm and eliminates the error caused by the deflection of the arm under load as it travels through its workspace. The control system consists of a laser, a light sensitive position sensor, a Delta Tau servo controller, and other circuitry.

1. INTRODUCTION

The common practice in robot arm design has been to eliminate any deflection in the joints and arms by severely over-designing the arm and the joints. This means that in order to counter gravitational forces and the resulting deflections, the robot arm is designed so strongly that for the nominal loads, the arm will have negligible deflections. As a result, although a robot arm may be able to withstand very large loads, its published and useful load capacity is very small. As an example, a large hydraulic robot such as Cincinnati Milacron T3 has a load capacity of 100 kg, but may easily lift over one ton, granted with large deflections. Compared to humans, where an Olympian weight-lifter may lift several times his weight, a robot only lifts a fraction of its weight. This is a result of the required positional accuracy for an arm.

As an alternative, it is proposed to design an arm which is not severely over-designed, but may deflect, and then correct the position of the arm actively, in real time. This will result in a much lighter arm, with smaller power requirements and weight, but with larger capacity, yet with same positional accuracy. The intention is not to design a flexible arm, but a semi flexible arm, which deflects slightly under the load.

Most robotic manipulators have up to six degrees of freedom, three for positioning, and three for orientation. Although the configuration of robots may differ significantly, in most robots, the positioning is accomplished by two successive arms. The wrist joint is usually attached to the base, and the waist joints are usually attached to short links. As a result, they do not significantly contribute to the total deflection of the end effector. Thus, it is generally sufficient to control the deflections of two main links of the arm.

In this project, an arm was constructed with a length of 33 cm, which can deflect up to 1.5 cm under the given load of 5 N. To detect the error of the tip position in the link, a laser light source was attached to the proximal side of the link, projecting a straight line parallel to the undeflected link. A sensor was attached to the distal end of the link, which can detect the laser light's position relative to the link. This information is used to calculate the additional movement that at any instant is needed in the robot's joints to eliminate the error. The controller will then compensate for the deflection by driving the arm motor faster or slower to drive the error to zero. To de-couple the correctional movements of the arm from the laser movements, the laser source is attached to a separate small motor which runs simultaneously with the joint motor, except for the correctional movements. Thus, the differential movement between the two motors, one driving the laser source and one driving the link, will eliminate the positional error caused by the deflection of the link.

The laser source is an inexpensive low power visible range laser diode source with a collimator. The collimator is focused and slotted to create a narrow line of projected light. The sensor is a

Hamamatsu Position Sensitive Device (PSD) which generates a current proportional to the position of the light spot relative to its end electrodes. The sensor thus detects the positional error between the laser light and the link. The laser light and the arm are both driven by servomotors with encoder feedback. The controller is a Delta Tau Programmable Multi-Axis Controller (PMAC) board.

2. PAST RESEARCH

An effective tip deflection control will constantly feed back tip position to the controller. This can be done in different ways. One way is to anticipate tip motion by mathematical relation to known values such as actuator position. As an example, Kelemen and Bagchi successfully used position of motor shaft angle to predict the location of the flexible arm tip [1]. Though successful in the lab, this technique may not be sufficient in an environment where disturbances such as a sudden change in payload, unexpected torsion, or permanent deformation in the arm can render the arm model irrelevant. Alternatively, tip position has been measured with a strain gauge placed at the base of the arm, or more directly, with a linear CCD array camera or a laser beam.

Both strain gauges and direct measurement have been used effectively in flexible motion control [2]. However, direct measurement requires a free path for the passage of light or other references. This could be a problem in an environment where there is smoke or steam, or where close proximity to surroundings can obstruct light path. Despite the possible complications, direct tip measurement is much more reliable than estimating tip position based on measurements at the base [3]. Niku [4] used a master-slave, two arm set up to measure the tip deflection, where the position of the loaded arm was compared to the position of an unloaded arm, moving in parallel to each other. Lewis [5] also used a laser light mounted inside the arm link to measure errors.

Engineers have used various types of algorithms in controlling flexible members: stable factorization, fuzzy logic, PID control, H-infinity, adaptive gain scheduling, and others [2,3]. Based on a system's individual constraints and performance expectations, one technique may work better than another.

3. SET-UP & INSTRUMENTATION

3.1. Flexible Arm Structure

Figure 1 is an illustration of the semi flexible arm structure. Though a seemingly simple system, its motion control problem is used as a basis for analyzing flexible, multi-link robots. The arm moves in a plane against the force of gravity and deflects under the end mass load. The actuator is a servomotor with a 500 liner-per-revolution optical encoder and a 214:1 gear ratio. The arm is 33 cm long, is made of aluminum, and is designed to deflect about 1.5 cm under the maximum load.

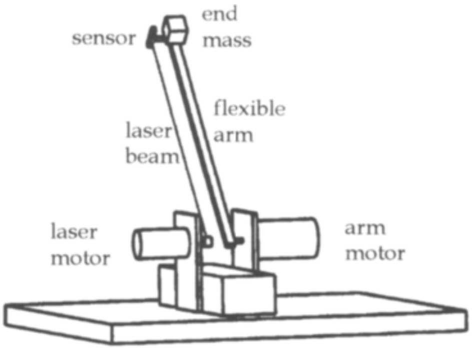


Figure 1: Flexible Arm Structure

The laser light is mounted on a bracket and is driven by a servomotor with a 187:1 gear ratio, and a 500 line-per-revolution optical encoder. The laser light source is an inexpensive, 5 mW, 770 nm, laser diode with a collimator which focuses the light on the sensor. The deflection of the arm is measured by a light sensor mounted at the distal end of the arm, normal to the laser light. The sensor used is a Hamamatsu Position Sensitive Device (PSD). When the laser is incident on the PSD, light energy is converted photoelectrically and is detected by two electrodes as photocurrent. The current in each half of the sensor is inversely proportional to the distance from each end electrode to the light spot (laser). The total amount of current generated by any given light, I_0, remains constant [6], while current at each end varies with the location of the light spot. Thus, the tip deflection of the arm can accurately be found by measuring the two currents.

3.2. PMAC & Accessories

A Programmable Multi-Axis Controller (PMAC) was used to control the system. PMAC plugs into the host computer bus to receive power and to communicate with the host computer. The PWIN1.7™ software program enables the user to configure, control, and trouble shoot the PMAC. It provides terminal interface - a text editor for writing programs, a watch window for viewing variables while a program is running, and a plotting buffer for gathering data.

The laser and the arm servomotors are programmed through PMAC software to move according to desired velocity profiles such as linear or s-curved. PMAC allows for manual or automatic tuning of the servomotors for best performance by choosing appropriate PID gains. This is referred to as the internal control loop. The external control loop is how PMAC is programmed to control tip deflection. The focus of this project is on the external control loop.

3.3. Feedback Circuitry

A circuit was designed and constructed to manipulate current signals from the sensor and to convert the analog signal to digital form. Figure 2 is a diagram of the feedback circuitry. A photocell was placed next to the light sensor to subtract ambient light. The final outcome of this circuit is a digital output, proportional to the position of the light spot on the sensor, which is fed back to PMAC.

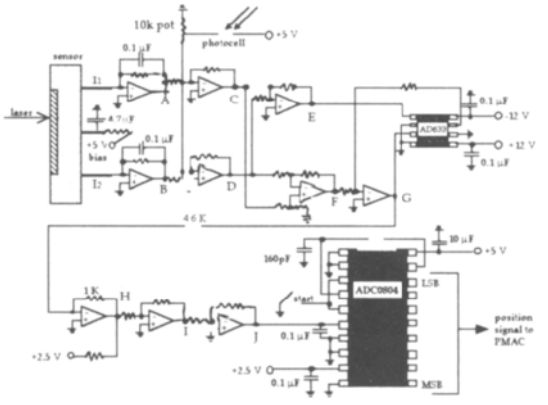

10k pot
photocell
+5 V
sensor
laser
0.1 µF
I1
I2
4.7 µF
+5 V bias
A
B
C
D
E
F
G
AD630
-12 V
+12 V
46 K
1K
H
I
J
160pF
10 µF
start
ADC0804
LSB
MSB
+2.5 V
position signal to PMAC

Figure 2: Feedback Circuitry and A/D Converter

Figure 3 illustrates the layout of instruments used in controlling the flexible arm.

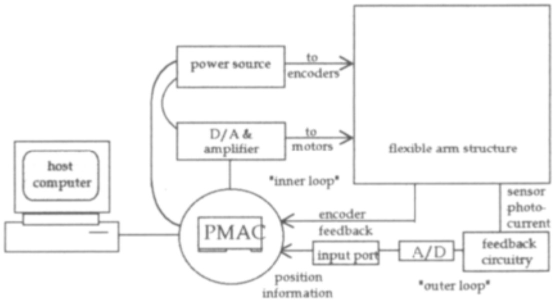

power source
to encoders
D/A & amplifier
to motors
flexible arm structure
host computer
"inner loop"
PMAC
encoder feedback
input port
A/D
feedback circuitry
sensor photo-current
position information
"outer loop"

Figure 3: General Layout of the system

4. THEORY

In modeling the arm it was assumed that the deflections are small enough such that linear relationships exist and that higher order natural frequencies will not be excited. The first natural frequency of the model arm was calculated and verified to be 4.1 Hz. Furthermore, it was assumed that the total mass of the beam was concentrated at the tip, since the weight of the beam compared to the point load at the tip was small.

Figure 4 shows a free-body diagram and mass-acceleration diagram used to find the differential equation of motion for the flexible arm. y is tip deflection from the "straight beam" shape. θ_{arm} is measured from the horizontal.

Summing moments about point O and using Newton's law to equate elements in the diagrams of Figure 4 gives the following differential equation of motion:

$$mL\frac{d^2y}{dt^2} + ky = mg\cos\theta_{arm} + mL^2\frac{d^2\theta_{arm}}{dt^2}$$

(1)

where m is mass of the end weight. Centripetal acceleration has been ignored since it is very small. It should be noted here that y is not the position error except when the arm is not corrected. Since the controller will be trying to correct the tip deflection error by compensating the rotation of the arm motor relative to the laser motor, in order to achieve zero error, it will increase θ_{arm}. As this happens, although y remains the same, error decreases. The following relationship exists between $\theta_{arm}, \theta_{laser}$, error, and y:

$$error = L(\theta_{laser} - \theta_{arm}) + y \qquad (2)$$

θ_{laser} is the desired position, error is the input to the controller, and θ_{arm} is the output. Equations (1) and (2) were linearized , transformed to the Z-plane, and used in the control loop.

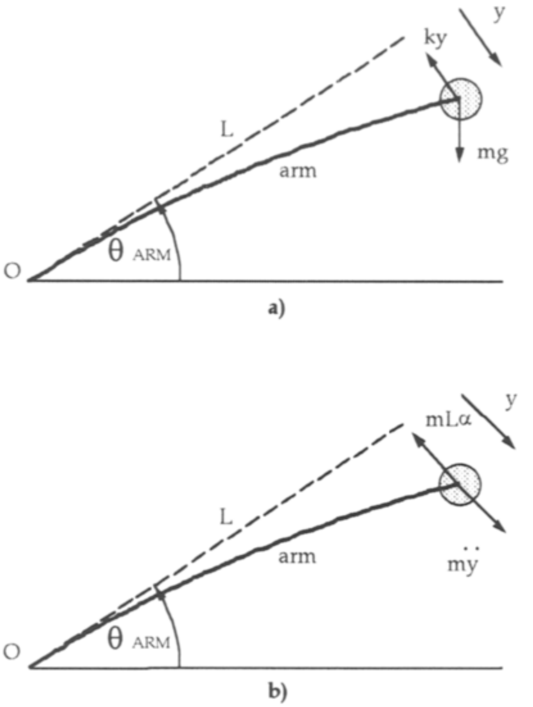


Figure 4: a) Free-Body and b) Mass-Acceleration Diagrams

5. SOFTWARE MODELING

To study the system's behavior, a computer model was generated and analyzed using Mathlab's Simulink™. Figure 5 is the control loop created in Simulink to model the semi flexible arm system. S_{arm} and S_{laser} are arc distances from horizontal.

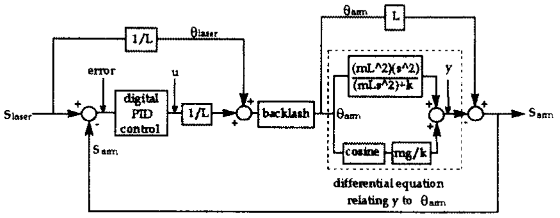


Figure 5: Simulink Control Model

Simulink can also model backlash. Since the gear box on the arm motor contains about 0.03 radians of backlash, it was modeled using this nonlinear library which is initialized by specifying the amount of play in a system. When the angle going into the box switches direction, the angle going out will not respond until the play is taken-up.

6. CONTROL PROGRAM

The dual motor scheme used for controlling the arm depends on two input variables: the angular position of the laser motor (input to the system) and the feedback signal (error) from the sensor. Laser and arm motors must be moved simultaneously, but at different values. To do this, a control program was used with two elements. As the motion element of the program moves the laser motor according to a chosen velocity profile, a second program monitors the error signal, stores it in memory, performs computational tasks, and actuates the arm motor by stepping it. In PMAC nomenclature, this is called a PLC program. The PLC program runs continuously in the background, from beginning to end, effectively in an infinite loop until it is told to stop.

Tuning the controller involves changing the PID coefficients and sampling rate to achieve a desired response. PMAC allows for auto-tuning of servomotors. This procedure was utilized to tune the controller, although the gain values were at time modified slightly to achieve better results depending on the aggressiveness of control and desired response. Ultimately, a dual control scheme with aggressive or mild control was used in conjunction with three laser motor velocity profiles (slow, medium, fast).

7. RESULTS

Figure 6 shows examples of the response of the arm (the positional error vs. time) for slow move with aggressive gains and for medium move with aggressive gains. Other situations resulted in similar results, with some exceptions. For example, with added mass to the arm, and with fast move and aggressive control, the arm could become unstable and out of control. There were also small movements at the end of the arm at rest caused by the backlash of the gearbox.

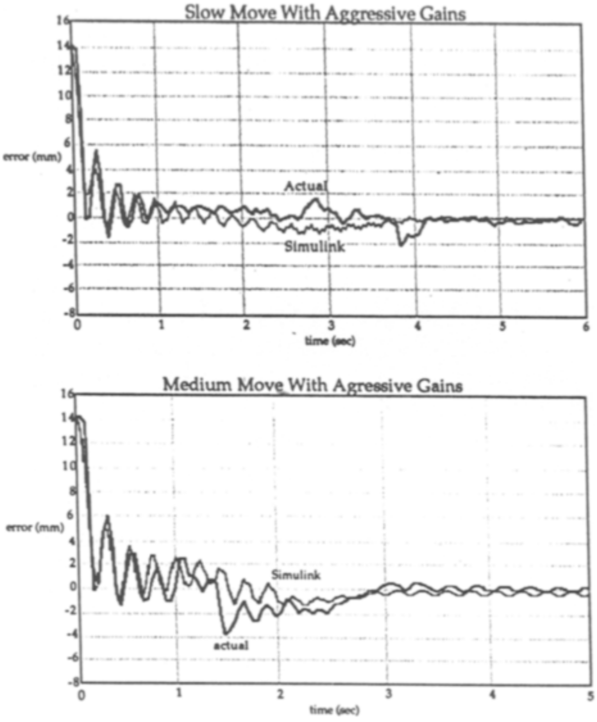


Figure 6: Arm response for slow move with aggressive gains and medium move with aggressive gains.

8. DISCUSSION

The discrepancy between the Simulink results and the actual results are due to simplifying assumptions. The model simulated by Simulink assumes linear system, point mass, mass concentrated on the neutral axis, etc. The real system is of course somewhat different and yields somewhat different results.

As can be seen, at the beginning of each cycle, there is a large deflection at the tip. The controller immediately tries to make the error zero. As the laser motor starts to move and the arm motor is commanded to follow, some vibration ensues which is kept to a minimum (±2 mm or less). After the arm stops, the error is kept at zero.

The system was capable of running the laser motor and the arm very fast while maintaining the error to a minimum, and without overshoot. The arm had no external damping added to it. This is an area where the design can be improved. For example, if the arm is made of materials which are either inherently more viscous, or if a sheath of material with high damping coefficient is wrapped around the arm, its behavior will improve.

This project demonstrates that it is possible to design an arm with less rigidity and at lower weights which can still be accurate. To demonstrate the possibility, in this project the arm was made much more flexible than needed. However, in a real robot, the arm can be less over-designed to save weight and make the robot more agile, and yet, maintain the accuracy needed.

REFERENCES

1. Keleman, Matei and Bagchi, Arunabha, 1992, "Modeling and Feedback Control of a Flexible Arm of a Robot for Prescribed Frequency-domain Tolerances," *Computers & Structures,* Vol. 39, No. 7, pp. 899-908.

2. Book, Wayne J., 1993, "Controlled Motion In an Elastic World," *ASME Journal of Dynamic Systems, Measurement, and Control,*, Vol. 115. pp. 252-260.

3. Cetinkunt, S. and Wu, S., 1990, "Tip Position of a Flexible One-Arm Robot With Predictive Adaptive Output Feedback Implemented With Lattice Filter Parameter Identifier," *Computers & Structures*, Vol. 36, No. 3, pp. 429-441.

4. Niku, Saeed B., "Scheme for Active Positional Correction of Robot Arms", *CAD/CAM, Robotics, and Factories of the Fututre '90,* Dwidevi, Verma, Sneckenberger editors, Springer Verlag, 1991, pp. 590-593.

5. Lewis, Jeremy, "The Development of an Optical End-Point Position Control System for Long Reach, Flexible Link Robots" http://www.mdx.ac/uk/www/ammc/jeremy.html.

6. *Hamamatsu Technical Data, Position-Sensitive Detectors.* Bridgewater, NJ, 1993.

7. Dabney, James B. and Harman, Thomas L.. *Simulink, Dynamic System Simulation for MATLAB.* Upper Saddle River, New Jersey: Prentice Hall, 1998, p.60.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

The Control of Elastic Multi-Link Manipulator Based on the Dynamic Compensation Method

Victor A. Utkin

Institute of Control Sciences, Profsoyuznaya 65, 117806, Moscow, Russia,
E-mail: vicutkin@ipu.rssi.ru

The paper presents a new approach to the flexible structure systems synthesis on the base of using a method to dynamic compensation and a division of motion on rate within the framework of systems with large coefficients and discontinuous control, allowing to decompose a problem of large dimensionality to the nonidependently solved a problems of smaller dimensionality.

1. INTRODUCTION

The paper presents a new approach to the design of rigid-flexible structure systems, based on a sliding mode technique [1] and on the method of dynamic compensation [2]. It is known, that robustness and order reduction are the most important virtue of variable structure control (VSC) with sliding mode.

The structure of rigid-flexible system can be presented as two mutually connected subsystems: the first subsystem describes the behavior of rigid coordinates, the second subsystem describes the vibration of flexure coordinates (usually it is expected that flexible variable are described end amount modes) [3]. The common approach to the control problem of flexible system, both for a continuous control algorithms [4] and for VSC' algorithms [5] consists of solution of two problems: the problem of stabilization of rigid subsystem and the restriction of flexible variables excitation. In this paper we suggest the control strategy, which gives a possibility to solve the stabilization problem of flexure subsystem directly. Such possibility appears to account of introduction to sidebar of feedback dynamic element (dynamic compensator or observer), reflect structure to vibrations of flexible arms of robot-manipulator [7]. Greatly, that proposed approach allows realize decompomposition of syntheses problem of large dimensionality on subproblem of smaller dimensianality.

The organization of this paper is as follows. Section 2 provides the background for our work. It introduces the dynamic compensation method in the invariance system design. Section 3 gives the synthesis procedure for a flexible structure systems based on the motion separation method. Finally, the application to a elastic multi-link manipulators is discussed in Section 4 using these techniques.

2. DYNAMIC COMPENSATION METHOD

Consider a linear dynamic system, described by equations

$$\dot{\theta} = A\theta + B(u + \Lambda q), \dot{q} = Wq + B_1 u, \qquad (2.1)$$

where $\theta \in R^n, u \in R^m, q \in R^k$-are vectors of the states, controls and disturbances; A, B, B_1, W, Λ - a matrixes of consistent dimensions. To stay the problem to ensure the stabilization to zero the state vector with invariance to external disturbances.
Let us chose the controls as follows:

$$u = -\Lambda z + v, \dot{z} = Wz + B_1 u + Lv, \qquad (2.2)$$

where $z \in R^k$ is the state vector of compensator and $v \in R^m$ is the vector of new control. The motions of closed system (2.1) and (2.2) are described by equations

$$\dot{\theta} = A\theta + B\Lambda\varepsilon + Bv, \dot{\varepsilon} = W\varepsilon + Lv, \qquad (2.3)$$

where $\varepsilon = q - z$. The new control is chosen in the class of systems with discontinuous control or high gains

$$v = -M\mathrm{sign}(s) \text{ or } v = -ks, \; s = C\theta, \tag{2.4}$$

where $s \in R^m$ is sliding surface, C is $(m \times n)$ constant matrix, M and k are positive constants. Under sufficiently large M and k, the overall motion of system (2.3) and (2.4) is divided into two phases: at the first stage the state vector reaches the sliding surface (or fast motion), whereas at the second stage the sliding (or slow) mode occurs along this surface. According to the equivalent control method the equation of sliding modes (or slow motion) will be described as follows:

$$\begin{aligned} \dot{\theta} &= [A - B(CB)^{-1}CA]\theta, s = C\theta = 0, \\ \dot{\varepsilon} &= -L(CB)^{-1}CA\theta + (W - L\Lambda)\varepsilon. \end{aligned} \tag{2.5}$$

The characteristic polynomial of the diagonal cell matrix of the system (2.5) coincides with the product of the characteristic polynomials of diagonal elements. Under assume that the pair (A,B) is controllable and (without loss of generality) the pair (Λ, W) is observable by appropriate choice of matrices C and L it is possible to arbitrarily independently assign (n-m) nonzero eigenvalues of matrix $A - B(CB)^{-1}CA$ and k eigenvalues of matrix $W - L\Lambda$. Note that the system (2.5) is homogeneous and, hence, the its stability can be assured with the aid of linear control with finite gains [2].

Also, note that the amplitude of discontinuous control (M) that ensure the fulfillment of conditions of sliding mode occurrence will be nonzero. After finish of transient process the undesirable self-oscillations (named as the chattering phenomena) is occurred. On the other hand, the equivalent control is tended to zero with state variables is tended to zero. Hence, the amplitude of discontinuous control can be diminished to zero, also. Thus, it is possible to eliminate the chattering phenomena.

3. CONTROL SYNTHESIZE OF FLEXIBLE STRUCTURE. BASIC IDEA.

In common view a model of multi-link manipulator with flexible arms may be described by the equations

$$\begin{aligned} \dot{\theta}_1 &= \theta_2, \\ \dot{\theta}_2 &= F(\theta_1, \theta_2, q) + B(\theta_1, q)\tau, \\ \dot{q} &= W(\theta_1, \theta_2, q) + B_0(\theta_1, q)\tau, \\ y &= D(\theta_1, q), \end{aligned} \tag{3.1}$$

where $\theta_1, \theta_2 \in R^n$ are rigid-body coordinates, the vector of flexure coordinates $q \in R^k$ is modeled by finite number modes, F, W are vector-functions, $B, B_0, (\mathrm{rank} B = n, \mathrm{rank} B_0 \le n)$ are functional matrixes of control distribution, $y \in R^1$ is tip position.

The follows control three steps procedure is suggested for the solution of the stabilization task of the system (3.1).

1 step.

Let us introduce a nonsingular transformation of coordinates as follows:

$$\bar{\theta}_2 = \theta_2 + K_1\theta_1,$$

$$\bar{q} = q + M\theta_2, M = -B^{-1}B_0.$$

Thus, the equations about of new flexure coordinates isn't consisted the control variables and stability of own dynamic of variables θ_1 may be supply by choice of matrix K_1:

$$\begin{aligned} \dot{\theta}_1 &= -K_1\theta_1 + \bar{\theta}_2, \\ \dot{\bar{\theta}}_2 &= \bar{F}(\theta_1, \bar{\theta}_2, \bar{q}) + \bar{B}(\theta_1, \bar{\theta}_2, \bar{q})\tau, \\ \dot{\bar{q}} &= \bar{W}(\theta_1, \bar{\theta}_2, \bar{q}) + \bar{B}_0(\theta_1, \bar{\theta}_2, \bar{q}), \\ y &= \bar{D}(\theta_1, \bar{\theta}_2, \bar{q}), \end{aligned} \tag{3.2}$$

where $\bar{B}_0 = \bar{M}(\theta_1, \theta_2, \bar{q})\bar{F}(\theta_1, \theta_2, \bar{q})$ and $\bar{F}, \bar{B}, \bar{B}_0, \bar{W}, \bar{D}$ has been get after substitution $q = \bar{q} - M\theta_2$ into F, B, B_0, W, D.

2 step.

In suggestion that system (3.1) is controllable (or stabilizable) for stabilization of flexure variables, described by third equation (3.1)) the variables $\overline{\theta}_2$ are chosen as the function of variables $\overline{q}$:

$\overline{\theta}_2 = L(\overline{q})$ and $L(0) = 0$.

3 step.

Based on the theory of discontinuous control systems, let us introduce the sliding mode along the surface $s = \overline{\theta}_2 - L(\overline{q}) = 0$ by choose control as follows:

$$\tau_i = \begin{cases} M_i^+(\theta_1, \overline{\theta}_2, \overline{q}), s_i > 0, \\ M_i^-(\theta_1, \overline{\theta}_2, \overline{q}), s_i < 0, \end{cases}$$

where functions M_i^+, M_i^- are continuous, $s = (s_1, s_2, \ldots, s_n) \in R^n$. Methods of synthesis of sliding mode in robotics control problems are considered in [8].

Thus, the motion of system (3.1) in sliding regime (s=0) will be describe by equations:

$$\begin{aligned} &\dot{\theta}_1 = -K_1\theta_1 + L(\overline{q}), \\ &\dot{\overline{q}} = \overline{W}(\theta_1, L(\overline{q}), \overline{q}) + \overline{B}_0(\theta_1, L(\overline{q}), \overline{q}), \\ &\overline{\theta}_2 = L(\overline{q}), \\ &y = \overline{D}(\theta_1, L(\overline{q}), \overline{q}) \end{aligned} \qquad \textbf{(3.3)}$$

So far as by choose of matrix K_1 and vector-function $L(\overline{q})$ it is supply the stability of system (3.3), the problem of tracking tip position may be solved by determination of demanded position of rigid-body generalized coordinates.

As rule, it is need to measure the generalized coordinates in absolute coordinate system. In the case, then it is possible to measure the tip position, it is need to change first equation (3.3) on the equation, writing about variables y. Then, if a manipulator structure surplus, follows a part of generalized coordinates to refer to second system (3.3) with according follows procedure.

Let $\operatorname{rank}\{\frac{\partial D(\theta_1, q)}{\partial \theta_1}\} = l < n$. Then, according with the implicit function theorem, the vector θ_1 may be divided on the two subvectors $\theta_1^t = [(\theta_1')^t, (\theta_1'')^t], \theta_1' \in R^l, \theta_1'' \in R^{n-l}$ thus, that $\operatorname{rank}\{\frac{\partial D(\theta_1', \theta_2'', q)}{\partial \theta_1'}\} = l$ and, consequently, the vector θ_1' may be written as a certain function from y: $\theta_1' = D^+(y, \theta_1'', q)$. Write the system (3.1) about variables y, θ_1'':

$$\begin{aligned} &\dot{y} = \{\frac{\partial D}{\partial \theta_1'}\}\theta_2' + \{\frac{\partial D}{\partial \theta_1''}\}\theta_2'' + \{\frac{\partial D}{\partial q}\}\dot{q}, \\ &\dot{\theta}_2' = F_1(\theta_1', \theta_1'', \theta_2', \theta_2'', q) + B_1(\theta_1', \theta_1'', q)\tau, \\ &\dot{\theta}_1'' = \theta_2'', \\ &\dot{\theta}_2'' = F_2(\theta_1', \theta_1'', \theta_2', \theta_2'', q) + B_2(\theta_1'\theta_1'', q)\tau, \\ &\dot{q} = W(\theta_1', \theta_1'', \theta_2', \theta_2'', q) + B_0(\theta_1', \theta_1'', q)\tau, \end{aligned} \qquad \textbf{(3.4)}$$

The decomposition procedure of stabilization problem of system (3.4) can look as follows. Supposing the variables θ_2' in first subsystem as fictitious control, will place its equal

$$\begin{aligned} s_y = \theta_2' + \{{}^{\partial D}\!/_{\partial \theta_1'}\}^{-1}[\{{}^{\partial D}\!/_{\partial \theta_1''}\}\theta_2'' + \\ +\{{}^{\partial D}\!/_{\partial q}\}\dot{q} + K_y y] = 0. \end{aligned} \qquad \textbf{(3.5)}$$

Then, by suitable choosing of matrix K_y the first subsystem of (3.4) will be stability: $\dot{y} = -K_y y$.

On the second step of procedure for performing the correlation (3.5) the sliding mode is synthesize along the surface $s_y = 0$.

Notice that for making the sliding motion is use by l component of control vector. Staying controls are synthesized on the scheme, stated above for system (3.1). Besides, if components of vector θ_2'' insufficiently for satisfactory stabilization latest subsystem (3.4), then in surface (3.5) follows add certain function of flexible variable (q).

4. APPLICATION TO MANIPULATOR CONTROL

4.1. Modeling of Flexible manipulators.

The dynamic of a manipulators with elastic arm can be written in the form

$$H(x)\ddot{x} = h(x,\dot{x}) + B\tau, \qquad \textbf{(4.1)}$$

where $x^t = (\theta^t, q^t)$ – is vector of generalized rigid-body coordinates $\theta \in R^n$ and flexure (elastic) coordinates $q \in R^k$, $H(x)$ is the positive definite mass matrix, $\tau \in R^n$ is the vector of joints torques, B is the control distribution matrix, $h(x,\dot{x})$ is the vector of generalized forces arising from centrifugal, coriolis, gravity effect and it associated with the elastic coordinates.

$$\begin{bmatrix} H_{11}H_{12} \\ H_{21}H_{22} \end{bmatrix}\begin{bmatrix} \ddot{\theta} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} h_1(x,\dot{x}) \\ h_2(x,\dot{x}) \end{bmatrix} + \begin{bmatrix} \tau \\ Wq \end{bmatrix}, \qquad \textbf{(4.2)}$$

where the matrix W corresponds to the natural vibration modes, provided that all actuators are locked [2]. The system (4.2) can be divided into two subsystems described the behavior of rigid-body and elastic coordinates

$$\ddot{\theta} = H_1^{-1}[H_{12}H_{22}^{-1}(h_2 - Wq) - (h_1 - \tau)], \quad \textbf{(4.3)}$$

$$\ddot{q} = H_2^{-1}[H_{21}H_{11}^{-1}(h_1 - \tau) - (h_2 - Wq)], \quad \textbf{(4.4)}$$

where $H_1 = H_{11} - H_{12}H_{22}^{-1}H_{21}$,

$$H_2 = H_{22} - H_{21}H_{11}^{-1}H_{12}.$$

The model (4.3) and (4.4) is used for control design. In the next subsection a state vector evolution problem is considered. The control scheme based on the sliding mode technique is prepared in subsection 4.3.

4.2. The observation problem.

Linear case.

Consider the observation problem of flexure coordinates $q \in R^n$ for the linear system

$$\dot{\theta} = A\theta + B(u + \Lambda q), \dot{q} = Wq + B_1 u, \qquad \textbf{(2.1)}$$

where vector $\theta \in R^n$ is available for measurements and pair $(W, B\Lambda)$ is observable.

Standard procedure of dynamic compensator (observer) syntheses consists in following.

Nonsingular change variable is interred $\overline{q} = q + L\theta$ and the second subsystem (2.1) is wrote about new coordinate:

$$\dot{\overline{q}} = (W + LB\Lambda)\overline{q} + (LA - WL - LB\Lambda L)\theta$$
$$+(LB + B_1)u.$$

Since pair $(W, B\Lambda)$ is observable, then matrix L exists so, that matrix $(W + LB\Lambda)$ has given distribution own number. Then dynamic device of similar structure

$$\dot{\overline{z}} = (W + LB\Lambda)\overline{z} + (LA - WL - LB\Lambda L)\theta$$
$$+(LB + B_1)u.$$

allows to get evaluation variable $\overline{q}$ on the strength of system stability, writing comparatively mismatches $\varepsilon = \overline{q} - z$: $\dot{\varepsilon} = (W + LB\Lambda)\varepsilon$.

Nonlinear case.

Now consider the observer design problem concerning the evolution of flexure coordinate q in the system (4.3) and (4.4) under assumption that variable θ and $\dot{\theta}$ are measurable. Also we use a simplified model in which the nonlinear terms do not depend from coordinate q and $\dot{q}$ [4]. Using nonsingular change of variable

$$q_1 = q + L_1(\theta)\dot{\theta}, q_2 = \dot{q} + L_2(\theta)\dot{\theta}$$

the system (4.3) can be write about new variable as follows

$$\dot{\overline{q}}_1 = \overline{q}_2 - L_1\overline{H}_1 W\overline{q}_1 + \overline{h}_1,$$
$$\dot{\overline{q}}_2 = (\overline{H}_2 - L_2\overline{H}_1)W\overline{q}_1 + \overline{h}_2, \qquad \textbf{(4.5)}$$

where

$$\overline{H}_1 = H_1^{-1}H_{12}H_{22}^{-1}, \overline{H}_2 = H_2^{-1}H_{21}H_{11}^{-1}$$

$$\overline{h}_1 = L_1\overline{H}_1(L_1\dot{\theta} + h_2) + L_1H_1^{-1}(h_1 + \tau) +$$
$$+\dot{L}_1\dot{\theta} - L_2\dot{\theta},$$

$$\overline{h}_2 = \overline{H}_2(h_1 - \tau) - H_2^{-1}h_2 + L_2\overline{H}_1h_2 + \\ +L_2H_1^{-1}(\tau - h_1) + \dot{L}_2\dot{\theta},$$

Describe the dynamic observer by equations

$$\dot{z}_1 = z_2 - L_1\overline{H}_1Wz_1 + \overline{h}_1,$$

$$\dot{z}_2 = (\overline{H}_2 - L_2\overline{H}_1)Wz_1 + \overline{h}_2 \qquad \textbf{(4.6)}$$

and write equation (4.5) and (4.6) in the mismatches $\varepsilon_i = \overline{q}_i - z_i, i = 1,2.$:

$$\dot{\varepsilon}_1 = \varepsilon_2 - L_1(\theta)D(\theta)\varepsilon_1,$$

$$\dot{\varepsilon}_2 = [\overline{H}_2W - L_2(\theta)D(\theta)]\varepsilon_1. \qquad \textbf{(4.7)}$$

where $D(\theta) = \overline{H}_1W$.

Consider the stabilization problem of system (4.7) sing a suitable choice of function matrixes $L_1(\theta)$ and $L_2(\theta)$ on the basis of the pole assignment approach under assumption that the system (4.7) for $L_1 = L_2 = 0$ is observable with respect to the output variable $y = D\varepsilon_1$ under any manipulator configuration. This is equivalent to the fact that the pair ($\overline{H}_2W, D$) is observable. Also, it is suggestion that follows condition is holds $rank(\overline{H}_2W) = rank(\overline{H}_1W) = rank(D) = k$ for any θ.

Consider master observer with constant matrixes, build, on example, for fixed deskside of manipulator:

$$\dot{\overline{\varepsilon}}_1 = \overline{\varepsilon}_2 - L_{01}D_0\overline{\varepsilon}_1, \\ \dot{\overline{\varepsilon}}_2 = (D_{01} - L_{02}D_0)\overline{\varepsilon}_1 \qquad \textbf{(4.8)}$$

where $\text{rank}(D_0) = n$,

$$\text{rank}(D_{01}) = \text{rank}(\overline{H}_2W) = \text{rank}(D).$$

Then it is readily verified that by choice of matrixes $L_1(\theta)$ and $L_2(\theta)$ according to relationships

$$L_1 = L_{01}D_0(DD^t)^{-1}D^t,$$

$$L_2 = (L_{02}D_0 - D_{01} + \overline{H}_2W)(DD^t)^{-1}D^t$$

the system (4.7) is reduced to the system (4.8).

Thus the variables $\overline{z}_i, i = 1,2$ can be given as estimates of variables $\overline{q}_1$ and $\overline{q}_2$ and, consequently, we have solved problem of reconstruction of variables q and $\dot{q}$.

4.3. Control design

In this subsection the approach presented in the section 3 is demonstrated for linear systems.

Linear case.

The idea of control procedure is demonstrated for linear time-invariant system described by equations

$$\dot{\theta}_1 = \theta_2, \dot{\theta}_2 = A_{21}\theta_1 + A_{22}\theta_2 + Bu, \\ \dot{q} = \dot{W}q + B_1u, \qquad \textbf{(4.8)}$$

where $\theta_1, \theta_2 \in R^n$ are vectors of position and velocity of rigid-body coordinates consistently, $q \in R^k$ is vector of flexible coordinates. The following procedure decomposes the control problem into three independent subproblems of smaller dimension.

On the first step is interred nonsingular change of flexible variable in the form

$$\overline{q} = q + M_1\theta_1 + M_2\theta_2,$$

where $M_2 = -B_0B^{-1}$, $M_1 = WM_2 - M_2A_{22}$.

Then the three subsystem (4.8) is written as follows

$$\dot{\overline{q}} = W\overline{q} + M\theta_1, \qquad \textbf{(4.9)}$$

where $M = M_2A_{21} - WM_1$. Since the system (4.9) is controllable about the vector θ_1, regarded as fictitious control there exists a feedback matrix F such that the matrix W+MF has any desired eigenvalue distribution. Thereby, the problem to stabilization of system (4.9) can be solved by means of choice of matrix F in the correlation

$$\theta_1 = F\overline{q} \qquad \textbf{(4.10)}$$

For ensuring an equality (4.10) on the second step of the procedure it is necessary to solve the stabilization problem of variable $\overline{\theta}_1 = \theta_1 - F\overline{q}$, described on the strength of (4.8) and (4.9) by equations

$$\dot{\overline{\theta}}_1 = (FW + MF)\overline{q} + M\overline{\theta}_1 + \theta_2 \qquad \textbf{(4.11)}$$

Will be considered variables in equation (4.11) as fictitious control and place its equal

$$\theta_2 = -(MF + FM)\overline{q}) - M\overline{\theta}_2 + K\overline{\theta}_1. \qquad \textbf{(4.12)}$$

Then, after substitution of this expression in (4.11) its stabilization is ensured by suitable choice matrix K:

$$\dot{\bar{\theta}}_1 = K\bar{\theta}_1.$$

Finally, on the third step of procedure necessary to ensure an equality (4.12) or convergence to zero the variables $\bar{\theta}_2 = \theta_2 + (MF + FM)\bar{q} + (M - K)\bar{\theta}_1$, which comply with equation of type

$$\dot{\bar{\theta}}_2 = \bar{A}_{12}\bar{\theta}_1 + \bar{A}_{22}\bar{\theta}_2 + \bar{W}\bar{q} + B_2 u. \qquad \textbf{(4.13)}$$

The problem to stabilization the system (4.13) can be solved by the choice of control u in the class of discontinuous feedback (similarly to second step) or in the class of discontinuous functions by means of introduction of sliding mode on the surface $s = \bar{\theta}_2 = 0$. Known that in discontinuous systems is ensured invariance to parameters of control object on the strength of that, that conditions of existence of sliding motion be the form of inequalities and, consequently, are to be provided under the known range of changing these parameters. In particular, this fact allows synthesize a stabilization problem for nonlinear system (4.13) in the suggestion, that are known bounds of nonlinear members, that inherent robot models.

Note that in practice realize moments (u) created by executive devices in class of discontinuous functions is not presented possible. In the case a model of control object must be complement by equations own dynamics of executive devices and intimate above hierarchical procedure must be continue up to real discontinuous controls [8]. Greatly note that specified decomposition is possible within the framework of discontinuous systems and large coefficients and, besides, is invariant to changing the parameters of control object and external disturbances.

5. CONCLUSION

In the paper is explored control problem of robot-manipulator with flexible arms. On the base of method to dynamic compensation is offered new approach to deciding a put problem, allowing realize a stabilization of robot model subsystem, described behavior of flexible variable.

Present results on the syntheses both a linear flexible structure models and nonlinear models, inherent robot models with flexible arms.

Offered control algorithms allow realize a decomposition of source syntheses control problem of large dimansionality nonindependently decided subproblems of smaller dimansionality. As follows, the syntheses problems divide into three subproblems-the problem of stabilization of model of flexible variable and two-level control in the space rigid-body coordinates.

REFERENCES

1. Utkin V.I. 1992 Sliding modes in control and optimization. N Y Sprigner Verlag.
2. V. A. Utkin and V. I. Utkin (1983) Design of invariant systems by the method of separation of motion.-Automation and remote Control, vol. 44, No. 12, Part 1, pp. 1559-1566.
3. Bales M.J.,1978, Feedback Control of Flexible systems, IEEE Transaction on Automatic Control, vol. AC - 23, no. 4, pp. 673 - 679.
4. Dong Li, 1994, Nonlinear control design for tip position tracking of a Flexible manipulator arm. Int. J. Control, vol. 60, no.6, pp. 1065 - 1082..
5. Young K. and Ozguner U., 1990, Frequency shaped variable structure control. Proc. VSS'90, pp. 181-185, Sarajevo, Yugoslavia, 19-20 Mart.
6. Drakunov S.V., Izosimov D.B., Lukyanov A.G., Utkin V.A. and Utkin V.I. (1990) The block control principle, I. Automation and Remote Control, Vol. 51, No. 5, Part 1, pp. 601 - 609.
7. Utkin V.A. 1990 Method of separation of motions in observation problems. Automation and Remote Control, Vol. 44, No. 12, Part 1, pp. 300 - 308.
8. Utkin V.A. and Izosimov D.B., 1990, Robot - manipulators control based on the method of movements separation. Proc. VSS' 90, Sarajevo, Yugoslavia, March, pp. 86 - 97.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Drive Dimension and Path Preparation for Parallel Manipulators

T.L. Tran and G. Kehl

Institute of Control Technology for Machine Tools, University of Stuttgart,
Seidenstr. 36, D-70174 Stuttgart, Germany.
e-mail: tang-long.tran@isw.uni-stuttgart.de

The development of new machine tools with parallel kinematics is currently one of the main activities of many research institutes and industrial manufacturers. Parallel manipulators require a very demanding control system. This paper describes the drive dimension and discusses some of the control problems, paying special regard to the path preparation for machines with parallel kinematics. Due to the complex relationship between the movements of the machine axes and those of TCP (Tool-Centre-Point) in the Cartesian coordinate system, the motion control of such machines must be carefully analysed with respect to possible collisions and the dynamic limits of the axes.

1. INTRODUCTION

When the Variax and Hexapod machine tools of Gidding & Lewis and Ingersoll were introduced at the International Machine Tool Show (IMTS) in Chicago, in 1994, a heated discussion about new structures in machine tool engineering was initiated [1,2,3]. These structures, so-called parallel kinematics, are not comparable with the majority of traditional machine tools and industrial robots based on an open kinematic chain. In fact, parallel kinematics have been known for a long time. In 1965, Stewart introduced the use of a parallel structure for a flight simulator. Therefore, parallel link systems are often referred to as the *Stewart Platform.*

A parallel manipulator is a closed-loop-mechanism in which all the linking elements are connected to the moving platform via joints. Serial manipulators, on the other hand, e.g. industrial robots consist of a number of links and joints so that each link is connected to its two adjacent links via two joints except the first and the last links. These serial manipulators generally have a long reach, a large workspace as well as high dexterity and good manoeuvrability. However, the serial structure results in manipulators with low stiffness and poor positioning accuracy. The advantages of parallel structures in comparison with the serial ones are well known. Besides the high load-weight ratio and the high structural stiffness, the possibility of keeping the drive units fixed to the base is an important characteristic of parallel manipulators allowing linear direct drives to be used by machine tools.

A large number of publications dealing with parallel kinematics have appeared in recent years. The problem of determining the workspace of parallel manipulators has been addressed by many authors [4,5,6]. Other articles discuss the mechanical design and its kinematic analysis [7,8]. Furthermore, the latest developments in parallel mechanisms include issues such as dynamics, control and applications [9,10,11]. However, most of these works refer to the kinematics with varying the lengths of the linking elements to change the position and the orientation of the moving platform. Besides that, very few publications dealing with the path preparation required for parallel manipulators can be found.

Based on [12, 13], the purpose of this paper is to present a method of drive dimension and to discuss the problems with respect to the motion control for parallel manipulators with linking struts of constant length.

2. DRIVE DIMENSION

Generally, the gear ratio of parallel manipulators is highly non-linear. This characteristic results in an expensive constructive dimension. Hence, this task must be computed in advance and is supported by a simulating process. To dimension the drives for a parallel manipulator it is important to determine the kinematic description which contains the forward and the inverse transformation. Considering the control system, the kinematic transformation must fulfil the real time requirements.

As an exemplary machine, a parallel manipulator (Figure 1) for the positioning with 3 axes is used for investigationing the kinematic parameters.

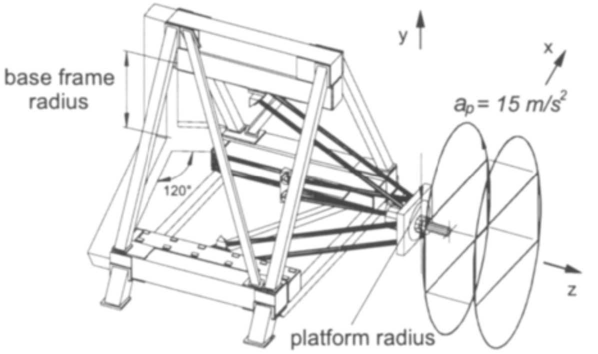


Figure 1. Parallel Manipulator

The linear drives are mounted on the base frame and connected to the platform via parallel strut pairs and ball joints. The platform carries the milling cutter which can be positioned relative to the fixed work piece. The orientation of the platform is independent from the software, but is fixed due to the use of strut parallelograms. As a result only translations of the platform are possible. The dimensions of the parallel manipulator are summerized in Table 1.

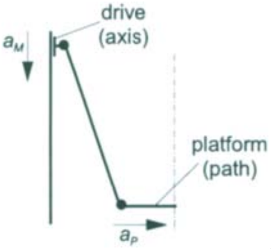
Table 1
Dimensions of exemplary Machine

geometry	
base frame radius	500 mm
base frame angle	120°
strut length	850 mm
distance of parallel struts	300 mm
platform radius	120 mm
motion	
workspace in x- and y-direction	600 x 600 mm
path acceleration in any direction in the whole workspace	15 m/s^2

For the depicted machine it is sufficient to investigate the kinematic properties in the xy-plane. The translation in z-direction only expands the workspace. The workspace is limited by the joint angles and the stroke of the linear drives.

The simulation results are shown on a map, which is based on isolines, connecting locations of identical properties. The simulated area is restricted to an area of 700x700 mm in the xy-plane. Grids characterize outer regions, which cannot be reached by the parallel manipulator.

The maximum axis acceleration required to achieve the desired path acceleration (a_P=15m/s^2) results directly from the maximum gear ratio i_{Max} between the three drives and the platform (Figure 2).



$$i = \frac{a_M}{a_P}$$

Figure 2. Gear Ratio

For platform motions in the xy-plane, the maximum gear ratio in the centre of the workspace amounts to only 0.51 and increases by 3-times in outer regions (Figure 3).

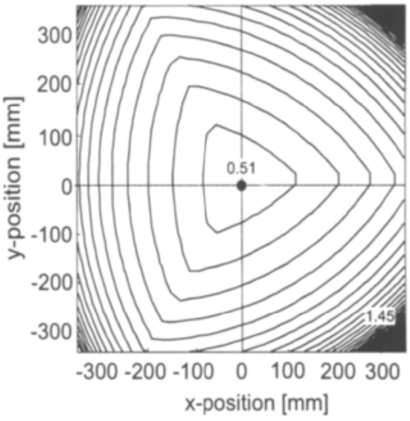


Figure 3. Maximum Gear Ratio in xy-Plane

With the proviso that the path acceleration has to be reached in any direction, the acceleration in z-direction is the worst-case for essential portions of the workspace, because the gear ratio in z-direction is exactly 1 and exceeds the gear ratio in the xy-plane. As a result the depicted maximum axis acceleration in the centre of the workspace shows an extensive plateau and increases clearly at the border of the workspace (Figure 4).

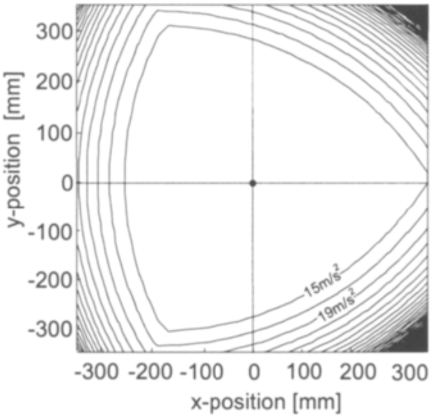


Figure 4. Maximum Axis Acceleration

The determined kinematic quantities are important inputs to the motion control, because the axis velocities and accelerations have to be reduced to their physical limits.

3. MOTION CONTROL

One of the most important tasks of the control system is the motion control, as shown in Figure 5.

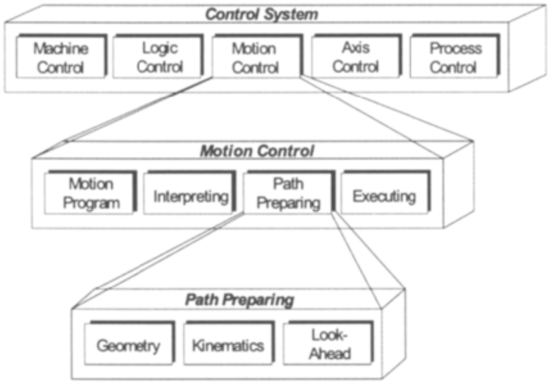


Figure 5. Structure of a Control System

Besides the determination and modification of the path, the motion control must also generate an optimal feed rate for a given trajectory.

Due to the complexity of this task, this chapter only focuses on problems of dynamic monitoring and collision avoidance.

3.1. Dynamic Monitoring

In general, the geometry of the path is defined by a NC programme. The timing of relative movements between the tool and the work piece is then calculated in the NC. From this it follows that there are certain boundary conditions, which have to be considered for each axis during the path preparation:

$$\left|\frac{dx_M}{dt}\right| \le v_{Max}, \quad \left|\frac{d^2x_M}{dt^2}\right| \le a_{Max}, \quad \left|\frac{d^3x_M}{dt^3}\right| \le j_{Max} \tag{1}$$

While the limits for the velocity and the acceleration are taken account of by the control system, the jerk limits are often ignored, particularly in the field of robotics. However, the actual jerk of the axis also determines the physic limits for the drives and is related to machine vibrations. If the programmed path is known, the relationship between the path velocity and the axis velocity can be expressed as follows:

$$\frac{dx_M}{dt} = \frac{dx_M}{ds} \cdot \frac{ds}{dt} = \frac{dx_M}{ds} \cdot v_P \tag{2}$$

Consequently, the velocity of each axis results from the path velocity v_P and dx_M/ds. The acceleration and the jerk of the axis can be obtained by differentiating further:

$$\frac{d^2x_M}{dt^2} = \frac{dx_M}{ds} \cdot a_P + \frac{d^2x_M}{ds^2} \cdot v_P^2 \tag{3}$$

$$\frac{d^3x_M}{dt^3} = \frac{dx_M}{ds} \cdot j_P + 3\frac{d^2x_M}{ds^2} v_P a_P + \frac{d^3x_M}{ds^3} v_P^3 \tag{4}$$

From equation (3), the path acceleration determines which axis takes part in the movement which changes the position and the orientation of the tool. Furthermore, the geometry of the path generally causes the axes acceleration to vary even when the path's velocity is constant. Therefore, during the motion control, the velocity, the acceleration and the jerk of each axis must be determined in order to prevent the axes from being dynamically overloaded.

The path preparation of a industrial control for conventional machine tools can be divided into the following steps:

1. Segmentation of the TCP path
2. Determination of the relationship between the axes movements and the path length
3. Calculation of a velocity profile
4. Generation the feed rate

The path is often segmented linearly. This is a oversimplification because the curvature of the path also influences the desired acceleration. Other segmentation methods depend on the curve length of the trajectory. The disadvantage of this method is the amount of computation involved. The segmentation can now be used for the path preparation as well as for the interpolation depending on the control system's method of interpolation.
Generally, there are two known variants:

- path interpolation
- axes interpolation

The main difference between the two interpolation methods is in the sequence of the data flow within the control system. In case of path interpolation, the kinematic transformation is computed after the path preparation and the polynomial interpolation. With axes interpolation, the results of the kinematic transformation are needed for the path preparation. The benefit of the first solution is the achievable accuracy of the path. On the other hand, with the second solution the movements of the axes are known before interpolation which simplifies their dynamic monitoring.

After the segmentation, it must be tested whether the dynamics of the axes lie within their previously mentioned limits. This task is also known as the look-ahead function. Hence, the main purpose is to generate a velocity profile over several path segments in order to follow a desired path with a optimal speed without exceeding the kinematic or dynamic limits of the machine and the drive units.
If the exact movements of the axes are not known before the interpolation, particular bases of the path must be transformed in advance in which case the desired machine co-ordinates are often obtained by approximating them with cubic or quintic splines. With the help of these splines it is possible to calculate the maximum acceleration and jerk for the path. Due to the path parameters dependency on time, the determined values are only significant for a discrete point and not for a whole segment. Nevertheless, these values are used as maximum limits for the path preparation. Furthermore the use of cubic splines does not allow an exact analysis of the jerk of the axes due to the increasing deviation of this approximation with higher derivatives. Therefore, in order to ensure the dynamic limits of the axes are not exceeded, the path limits must be respected during the path preparation.

In case of the axes interpolation the movements of the axes are determined at path preparation. The machine co-ordinates are obtained after the kinematic transformation and are often described with splines of fifth order. In comparison with cubic splines, this technique has a better convergence behaviour and interpolating characteristics. Subsequently, the maximum path velocity can be computed analytically from the given splines for in the given boundary conditions. However, this determination of the necessary path parameters involves various linearisations needed in order to simplify and to reduce the required computing operations. For example, it is assumed in [14] that the movements of the axes are lightly curved. The existing acceleration and jerk of the axes caused by a circular trajectory cannot, however, be obtained with the help of such simplifications (Figure 6).

Currently, industrial 5-axis machine tools consist of two kinematic chains: 3 linear axes for positioning and two rotary axes for the orientation. In most cases, the rotary axes are ignored during the path preparation. For this reason, a Cartesian system can be used for determining the path parameters. Hence, the linearisations involved in the above mentioned methods for motion control deliver acceptable results.

On the other hand, parallel manipulators cannot be designed with a Cartesian configuration. This means that the movements of the axes have a non-linear coherence to the given trajectory. Therefore, the path preparation for parallel manipulators has high requirements with respect to determining it's dynamic characteristics and the avoidance of overloading of the drive units. High dynamic drive units require a jerk limiting. Due to the mentioned non-linear relationship, it must be considered that the jerk of each axis must be calculated with sufficient accuracy during the whole path preparation. Besides that, it is also known that dynamic overloads can still appear for 5-axis manufacturing even if the path acceleration is limited to the weakest axis of the machine caused by extreme changes of the orientation. In addition, the different band widths between the path and the axes as a result of the non-linear transformation have the consequence that the advantages of the acceleration curve with a function $\sin^2$ cannot be guaranteed. Considering the development work and the computing cost, a jerk limited slope with a trapezoid form for the acceleration is therefore suggested.

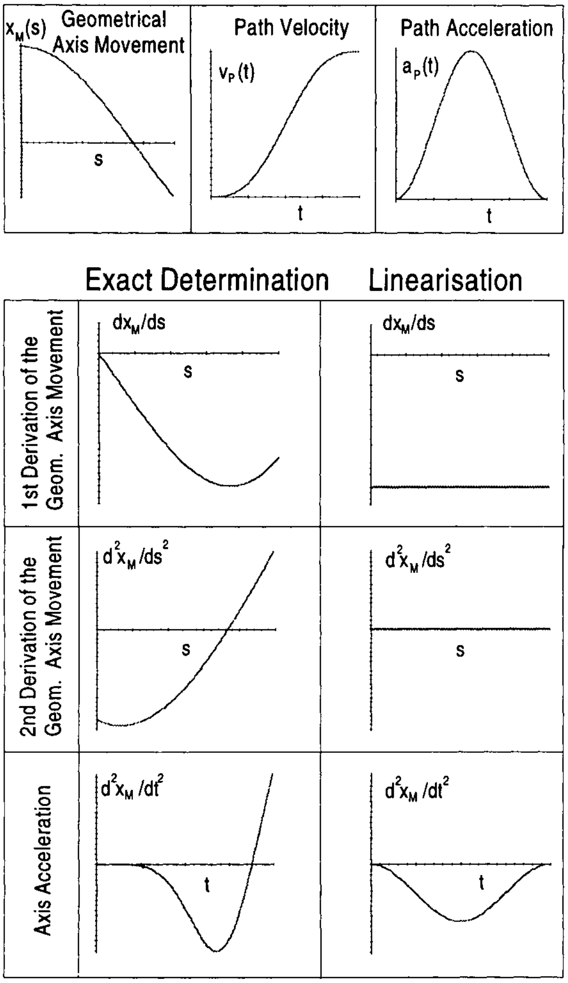


Figure 6. Different Determinations of the Axis Acceleration

In fact, the path preparation for parallel manipulators involves problems similar to those found in robotics. But since most robots are used for transport and assembly and the path is purely determined by a start point and a end point, there are no high requirements for path accuracy. Other applications with demanding path preparation have off-line solutions for generating the velocity profile. But this solution is not suitable for parallel manipulators when applied as machine tools. Due to the described problems, the current control systems applied for parallel manipulators reduce the real dynamic characteristics of the axes for the path preparation to ensure the safety of the drive units. However, this feature means that the dynamic capability of the drives cannot be exhausted [15].

3.2. Collision Avoidance

Another important task of the path preparation is the consideration of possible collisions. To solve this problem for conventional machine tools, it is normally simply determined using the software limit switches for each axis. In this way, the area of the axis movements is fixed and therefore the workspace of the machine.

Due to the large possibilities of the axes and the complex interaction of the moving parts, the problem of collision must be analysed with a special regard for parallel kinematics. Even an obvious collision between the moving platform with the base is not comparable to the existing machine tools. In this case, the software limit switches must be determined by including the transformation equations. This is necessary due to the need for all axes to describe the movements of the TCP in a layer. Besides that, the determination of the permitted workspace is also essential to avoid collision between the moving platform and the machine frame.

A further characteristic for parallel manipulators is the consideration of the moving struts. These can cause unwished collisions although the TCP is placed in the permitted workspace. In some extreme positions of the TCP, it is possible that the linking struts touch the machine frame. In addition, special kinematics e.g. shown in Figure 6 are the reason for further collisions. The manipulator has a moving platform with two levels to which the struts are connected. By changing the orientation of the TCP, the possible collision between the bottom struts and the upper level and vice versa has to be avoided. The determination of dangerous positions needs a very high computing effort. Therefore, the study of the permitted workspace can be performed in advance to ensure a safe operating mode. Finally, the computed software limit switches can be put into the control system and used for the path preparation.

Furthermore, Figure 7 shows the possibility for parallel manipulators to use two linear direct drive units on one drive sledge. In these cases, the determination of software limit switches is not sufficient anymore. Consequently, the distance between the two drives must also be watched. By using splines to describe the axis movements, the positions of the axes are obtained as follows:

$$x_M(\chi) = a_0 + a_1\chi + a_2\chi^2 + a_3\chi^3 + \ldots \qquad (5)$$

$$y_M(\chi) = b_0 + b_1\chi + b_2\chi^2 + b_3\chi^3 + \ldots \qquad (6)$$

The polynomial equation for the distance can then be set up in the form:

$$\begin{aligned} d(\chi) &= (a_0 - b_0) + (a_1 - b_1)\chi \\ &\quad + (a_2 - b_2)\chi^2 + (a_3 - b_3)\chi^3 \ldots \end{aligned} \qquad (7)$$

Finally, the requirement

$$d(\chi) \geq \Delta_{Min} \qquad (8)$$

must also be considered during the path preparation.

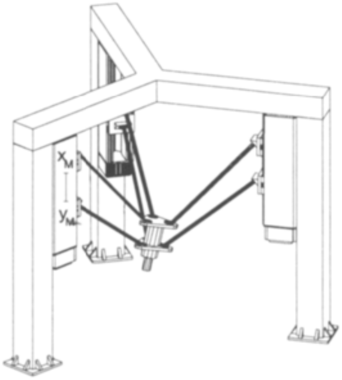


Figure 7. Parallel Manipulator for 5-axis Manufacturing

4. CONCLUSION

This paper has described a method to dimension the drives described the problems during the path preparation for parallel manipulators. Specifically, the motion control of conventional control systems is compared and analysed with respect to the characteristics of parallel kinematics. The issue of collision avoidance is also discussed.

REFERENCES

1. METALLWORKING Engineering and Marketing: Revolution in machining center structure, Second Quarter 1995, pp. 27-29.
2. Neugebauer, R.; Wieland, F.: Neue Werkzeugmaschinenstrukturen, ZWF 91 (1996), Vol. 7-8, pp. 363-366.
3. VDI-Nachrichten: Steuerungstechnik verbessert den Werkzeugmaschinenbau, Nr. 39, 27.9.1996.
4. Merlet, J.-P.: Workspace-oriented methodology for designing a parallel manipulator, IEEE Int. Conf. On Robotics and Automation, pp. 3726-3731, Minneapolis, 24-26 April 1996.
5. Yang, P.-H.; Waldron, J.K.; Orin D.E.: Kinematics of a three degree-of-freedom motion platform for a low cost driving simulator, Recent Advances in Robot Kinematics, pp. 89-98, Kluwer, 1996.
6. Gosselin, C.M.; Lavoie; E.; Toutant, P.: An efficient algorithm for the graphical representation of the three-dimentional workspace of parallel manipulators, 22nd Biennial Mechanisms Conf., pp. 323-328, Scottsdale, 13-16 September 1992.
7. Raghavan, M.; Roth, B.: Solving polynomial systems for the kinematic analysis and synthesis of mechanisms and robot manipulators, Special 50th Anniversary Design Issue, Vol. 117 pp. 71-79, June 1995.
8. Innocenti, C; Parenti-Castelli, V.: Exhaustive enumeration of fully parallel kinematic chains, ASME., Dynamic Systems and Control, Vol. 55-2, pp. 1135-1141, 1994.
9. Zhuang, H.; Liu, L.: Self calibration of a class of parallel manipulators, IEEE Int. Conf. On Robotics and Automation, pp. 994-999, Minneapolis, 24-26 April 1996.
10. Nguyen, C.C. et al.: Adaptive control of a Stewart platform based manipulator, Journal of Robotic Systems, Vol. 10-51, pp. 657-687, July 1993.
11. Geng, Z; Haynes L.S.: Neural network for the forward kinematics problem of a Stewart Platform, IEEE Int. Conf. On Robotics and Automation, pp. 2650-2655, Sacramento, 11-14 April 1991.
12. Pritschow, G.; Wurst, K.-H.: Systematic design of hexapods and other parallel link systems, Annals of the CIRP, Vol. 46-1 pp. 291-295, 1997.
13. Pritschow, G.; Wurst, K.-H.: LINAPOD - Ein Baukastensystem für Stabkinematiken, wt Werkstattstechnik 87, pp. 437-440, Sept.-Oct. 1997.
14. Fauser, M.: Steuerungstechnische Maßnahmen für die Hochgeschwindigkeits-Bearbeitung, Dissertation RWTH Aachen, 1997.
15. Pritschow, G.; Tran, T.L.; Wildermuth, D.: Bahnvorbereitung für Werkzeugmaschinen mit paralleler Kinematik, Chemnitzer Parallelstruktur Seminar, Chemnitz, 28-29 April 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

PASSIVITY BASED ADAPTIVE CONTROLLER FOR A TWO DEGREES OF FREEDOM ROBOT USING A COMPOSITE PARAMETERS ADAPTATION LAW

Ştefan Dumbravă

Automatic Control & Industrial Informatics Department,
"Gh. Asachi" Technical University of Iaşi,
7-9 Horia Street, 6600, Iaşi, România

In this paper, a high accuracy control method of robot joint trajectory tracking, with the realistic assumption of uncertainties in model parameters, is studied. The method exploits the passivity property of the mechanical structure of the system, which is maintained in closed loop by a proper controller. The global asymptotic stability of the overall non-linear system and convergence of the estimated parameters to exact ones are proved. The method is demonstrated for a two degrees of freedom robot structure. The simulation results emphasise the theoretical conclusions and give a measure of the performances obtained in case of a practical implementation.

1. INTRODUCTION

Contrary to general theoretical methodologies of robots motion control, the passivity-based approach takes advantage of the robot system physical structure, using its passivity property, [1-4]. The notions of passivity and strict passivity on L_{2e} (the extension of n dimensional square integrable functions space), are defined conform [5].

Definition 1. Let $H: L_{2e} \to L_{2e}$. Then H is passive if there exists a constant β so that: $\langle Hx, x\rangle_T \geq \beta$, $\forall x \in L_{2e}$, $\forall T \geq 0$, while H is strictly passive if there exist a $\gamma > 0$ and a constant β so that: $\langle Hx, x\rangle_T \geq \gamma \|x\|_{2T}^2 + \beta$, $\forall x \in L_{2e}$, $\forall T \geq 0$.

These notions are used to prove the stability of the system. Considering the robot structure, the non-linear dynamic model with friction torques included it is used, (1).

$$\tau_e = M(q,\theta)\ddot{q} + C(q,\dot{q},\theta)\dot{q} + G(q,\theta) + F(\dot{q},\theta) \quad (1)$$

In (1), $q, \dot{q}, \ddot{q}$ are $(n\text{x}1)$ vectors representing the position, velocities and the accelerations of the robot joints; τ_e is the $(n\text{x}1)$ vector of external torques, $M(q,\theta)$ is the $(n\text{x}n)$ inertia matrix, $C(q,\dot{q},\theta)\dot{q}$ the $(n\text{x}1)$ vector of the Coriolis and centrifugal torques, $G(q,\theta)$ the $(n\text{x}1)$ vector of gravitational torques and $F(\dot{q},\theta)$ the $(n\text{x}1)$ vector of friction torques. In this paper, the friction torque is considered composed of Coulomb and viscous components. The friction torque is considered a symmetric function of $\dot{q}$. θ is a $(p\text{x}1)$ vector containing the parameters of the robot dynamic model. Using (1), there exists a reparametrization of all unknown parameters into a parameter vector $\theta \in R^p$ that enters linearly in the system model.

$$\tau_e = M_k(q)\ddot{q} + C_k(q,\dot{q})\dot{q} + G_k(q) + Y(q,\dot{q},\ddot{q})\theta \quad (2)$$

where: $M_k(.), C_k(.), G_k(.)$ represent the known parts of the system dynamics and $Y(q,\dot{q},\ddot{q})$ is the regressor matrix of dimension $(n\text{x}p)$. In equation (2), the friction coefficients are considered unknown and they are included in vector θ. The passivity-based controller is designed so that the passivity property of the closed loop system to be maintained together the control objective and the parameters adaptation law.

2. PASSIVITY BASED ADAPTIVE CONTROL

The adaptive controller proposed in [1], [6] and described by the equations (3-7), is considered. It will be completed by an adaptation law which ensures a quick convergence of the estimated parameters and uses as external information only the desired position and velocity . The condition related to the passivity of the overall system will be further proved.

$$\begin{aligned}\tau_e &= M_k(q)\ddot{q}_r + C_k(q,\dot{q})\dot{q}_r + G_k(q)\\ &+ Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)\hat{\theta} - \upsilon\end{aligned} \tag{3}$$

$$\dot{q}_r \equiv \dot{q}_d - \Lambda e \tag{4}$$

$$e = q - q_d \tag{5}$$

$$\upsilon = -K_d s_1 \tag{6}$$

$$s_1 \equiv \dot{q} - \dot{q}_r = \dot{e} + \Lambda e \tag{7}$$

In (3-7), $q_d, \dot{q}_d$ are the vectors of the desired position and velocity, and: $\Lambda > 0$, $K_d\ 0$ are diagonal matrices. From (2-7), the following error equation of the closed-loop system results:

$$M(q,\theta)\dot{s}_1 + C(q,\dot{q},\theta)s_1 + K_d s_1 = Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)\tilde{\theta} \tag{8}$$

where: $\tilde{\theta} = \hat{\theta} - \theta$. $\hat{\theta}$ is the vector of the estimated unknown constant parameters. It is denoted as:

$$\bar{\tau} = M(q,\theta)\dot{s}_1 + C(q,\dot{q},\theta)s_1 \tag{9}$$

In order to increase the speed of convergence of the estimated parameters, the gradient type adaptation law is augmented with a term function of the prediction error on the control torque. This *composite parameter adaptation law*, (10), is used together the controller equations (3-7).

$$\frac{d}{dt}\tilde{\theta}(t) = -\Gamma(t)[Y^T(q,\dot{q},\dot{q}_r,\ddot{q}_r)s_1 + V^T(q,\dot{q})W\varepsilon_f] \tag{10}$$

$V(q,\dot{q})$ is the filtered regressor matrix by a stable and proper low pass filter with transfer function:

$$H_F(s) = \frac{1}{Ts+1} \tag{11}$$

where: $T > 0$ some time constant. $W = W^T\ 0$ is a weighting diagonal constant matrix and ε_f is the difference between the filtered prediction torque, denoted as $\hat{\tau}_{ef}$, and the filtered actual control torque, denoted as τ_{ef}, by the same low pass filter with transfer function $H_F(s)$:

$$\varepsilon_f = \hat{\tau}_{ef} - \tau_{ef} \tag{12}$$

Since from (2) it results that the difference between the prediction torque and the actual control torque is:

$$\varepsilon = \hat{\tau}_e - \tau_e = Y(q,\dot{q},\ddot{q})\tilde{\theta} \tag{13}$$

consequently it follows that:

$$\varepsilon_f = \hat{\tau}_{ef} - \tau_{ef} = V(q,\dot{q})\tilde{\theta} \tag{14}$$

The requirement for filtering the prediction error is obvious from the equations (13-14). While the computation of the regressor matrix $Y(.)$ requires the acceleration vector, this is not necessary in (14), since $V(.)$ does not depend on it as a result of filtering operation. For generating the variable in time adaptation gain matrix, the *cushioned-floor method*, [6], is used, (15).

$$\begin{aligned}\frac{d}{dt}\Gamma(t) &= \alpha(t)[\Gamma(t) - \Gamma(t)K_0^{-1}\Gamma(t)] +\\ &-\Gamma(t)V^T(q,\dot{q})WV(q,\dot{q})\Gamma(t)\end{aligned} \tag{15}$$

where: $K_0 = K_0^T > 0$ is a constant diagonal matrix and $\alpha(t) \geq 0$, $\forall t \geq 0$, is called *the variable forgetting factor*. The initial value of the adaptation gain matrix, $\Gamma(0)$, is chosen as $\Gamma^{-1}(0) - K_0^{-1} \geq 0$, which is proved that implies: $\Gamma(t) \leq K_0$, whatever $t \geq 0$. The condition has a practical importance, imposing an upper bound of the adaptation gain matrix $\Gamma(t)$.

The system described by equations (8-10) is represented in Fig. 1, where H_1, H_2, H_3 are causal operators mapping L_{2e} into L_{2e}.

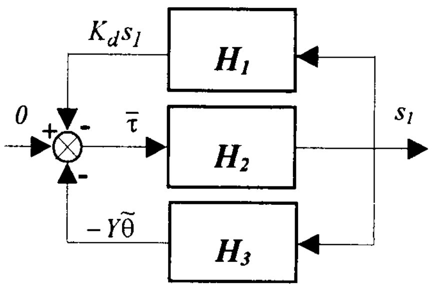


Figure 1. The block diagram for passivity analysis.

Property 1. The interconnected substructures of the system, Fig. 1, satisfy the following properties: H_1 is strictly passive, H_2 and H_3 are passive.
Proof:

$$\langle s_1, K_d s_1\rangle_T \geq \gamma_1 \|s_1\|_{2T}^2 \tag{16}$$

where it was denoted as: $\gamma_1 > 0$ the smallest eigenvalue of matrix K_d. Consequently H_1 is strictly passive.

$$\langle \bar{\tau}, s_1\rangle_T = \int_0^T s_1{}^T [M(q,\theta)\dot{s}_1 + C(q,\dot{q},\theta)s_1]dt$$

$$= \frac{1}{2}\int_0^T \frac{d}{dt}(s_1{}^T M(q,\theta)s_1)dt \geq -\frac{1}{2}s_1{}^T(0)M(q,\theta)s_1(0)$$

The property that: $\dot{M}(q,\theta) - 2C(q,\dot{q},\theta)$ is a skew matrix and consequently its quadratic form is null, was used. Then, it is chosen:

$$\beta_1 = -\frac{1}{2}s_1^T(0)M(q,\theta)s_1(0),\ \beta_1 \leq 0$$

and consequently:

$$\langle \bar{\tau}, s_1\rangle_T \geq \beta_1 \tag{17}$$

implying that H_2 is passive.

$$\left\langle s_1, -Y\tilde{\theta}\right\rangle_T = \int_0^T -s_1^T Y\tilde{\theta}dt$$

$$\dot{\tilde{\theta}} = -\Gamma Y^T s_1 - \Gamma V^T W V \varepsilon_f$$

$$-Y^T s_1 = \Gamma^{-1}\dot{\tilde{\theta}} + V^T W V\tilde{\theta}$$

$$\left\langle s_1, -Y\tilde{\theta}\right\rangle_T = \int_0^T (\tilde{\theta}\Gamma^{-1}\dot{\tilde{\theta}} + \tilde{\theta}^T V^T W V\tilde{\theta})dt$$

$$\frac{1}{2}\frac{d}{dt}(\tilde{\theta}^T\Gamma^{-1}\tilde{\theta}) = \tilde{\theta}^T\Gamma^{-1}\dot{\tilde{\theta}} + \frac{1}{2}\tilde{\theta}^T \frac{d}{dt}(\Gamma^{-1})\tilde{\theta}$$

$$= \tilde{\theta}^T\Gamma^{-1}\dot{\tilde{\theta}} + \frac{1}{2}\tilde{\theta}^T[-\alpha(t)[\Gamma^{-1} - K_0^{-1}] + V^T W V]\tilde{\theta}$$

$$\left\langle s_1, -Y\tilde{\theta}\right\rangle_T = \frac{1}{2}(\tilde{\theta}^T\Gamma^{-1}\tilde{\theta})\Big|_0^T$$

$$+\frac{1}{2}\int_0^T \alpha(t)\tilde{\theta}^T(\Gamma^{-1} - K_0^{-1})\tilde{\theta}dt + \frac{1}{2}\int_0^T \tilde{\theta}^T V^T W V\tilde{\theta}dt$$

Due to the fact that: $W=W^T>0$, the matrix Γ^{-1}-K_0^{-1} is symmetric and positive semidefinite, [7], and $\alpha(t)\geq 0$, the following lower bound can be found for the integrals:

$$\begin{aligned}&\frac{1}{2}\int_0^T \alpha(t)\tilde{\theta}^T(\Gamma^{-1} - K_0^{-1})\tilde{\theta}dt + \frac{1}{2}\int_0^T \tilde{\theta}^T V^T W V\tilde{\theta}dt\\ &\geq \frac{\delta_{\Gamma K_0}}{2}\|\tilde{\theta}\|_{2T}^2 + \frac{\delta_W}{2}\|V\tilde{\theta}\|_{2T}^2 \geq 0\end{aligned} \tag{18}$$

where $\delta_{\Gamma K_0}$, δ_W are the smallest eigenvalues of the matrices: (Γ^1-K_0^{-1}) and W, respectively. Thus:

$$\begin{aligned}&\left\langle s_1, -Y\tilde{\theta}\right\rangle_T \geq \frac{1}{2}\tilde{\theta}^T(T)\Gamma^{-1}\tilde{\theta}(T) - \frac{1}{2}\tilde{\theta}^T(0)\Gamma^{-1}\tilde{\theta}(0)\\ &+\frac{\delta_{\Gamma K_0}}{2}\|\tilde{\theta}\|_{2T}^2 + \frac{\delta_W}{2}\|V\tilde{\theta}\|_{2T}^2\end{aligned} \tag{19}$$

$$\beta_2 = -\frac{1}{2}\tilde{\theta}^T(0)\Gamma^{-1}(0)\tilde{\theta}(0)$$

and since Γ^{-1} is positive definite, it results (20) and the fact that H_3 is passive.

$$\left\langle s_1, -Y\tilde{\theta}\right\rangle_T \geq \beta_2 \tag{20}$$

The stability analysis of interconnected non-linear systems using passivity theory is further considered.
Theorem 1. Consider the feedback system in Fig. 1, where H_1, H_2, H_3 map L_{2e} into L_{2e}, and suppose that exist solutions in L_{2e}. Assume that H_1 is strictly passive, and H_2, H_3 are passive. Then $s_1 \in L_2$.
Proof: Using (8), the equation (9) can be written as:

$$\bar{\tau} + K_d s_1 = Y\tilde{\theta} \tag{21}$$

and consequently, equation (22). Then, using Property 1, it results (24):

$$\langle s_1, \bar{\tau}\rangle_T + \langle s_1, K_d s_1\rangle_T + \left\langle s_1, -Y\tilde{\theta}\right\rangle_T = 0 \tag{22}$$

$$\beta_1 + \gamma_1\|s_1\|_{2T}^2 + \beta_2 \leq 0 \tag{23}$$

$$\gamma_1\|s_1\|_{2T}^2 \leq -\beta_1 - \beta_2 \Rightarrow \|s_1\|_{2T} \leq \sqrt{\frac{-\beta_1 - \beta_2}{\gamma_1}} \tag{24}$$

Since β_1, β_2 and γ_1 are constants, (24) is true $\forall T \geq 0$, and that implies: $s_1 \in L_2$. ❖

Theorem 2. Let $H(s) \in R^{nxn}(s)$ be an exponentially stable and strictly proper transfer function, and let $y = H(s)u$. Then,

(i) if $u \in L_2$, then $y \in L_2 \cap L_\infty$, $\dot{y} \in L_2$, y is continuos, and $\lim_{t \to \infty} y(t) = 0$.

The proof is found in [5]. ❖

From (7), it results that:

$$e = (sI + \Lambda)^{-1} s_1 \tag{25}$$

and since $H(s) = (sI + \Lambda)^{-1}$ fulfils the conditions of Theorem 2, and $s_1 \in L_2$ (Theorem 1), from both it can be concluded, (Theorem 2,(i)) that $e \in L_2 \cap L_\infty$, $\dot{e} \in L_\infty$ and $\lim_{t \to \infty} e(t) = 0$. This implies global asymptotic convergence of the position tracking errors e. ❖

Remark 1. In the conditions of Theorem 2, $\lim_{t \to \infty} \tilde{\theta}(t) = 0$.

Proof: The relation (21) can be re-written as in (26):

$$\left\langle s_1, -Y\tilde{\theta} \right\rangle_T \geq \frac{\delta_m}{2} \tilde{\theta}^T(T)\tilde{\theta}(T) + \beta_2 \tag{26}$$

where: $\delta_m > 0$ is the smallest eigenvalue of Γ^{-1}. From (16), (17), (22) and (26), it is obtained (27):

$$\gamma_1 \|s_1\|_{2T}^2 + \beta_1 + \beta_2 + \frac{\delta_m}{2} \tilde{\theta}^T(T)\tilde{\theta}(T) \leq 0 \tag{27}$$

$$\beta_1 + \beta_2 + \frac{\delta_m}{2} \tilde{\theta}^T(T)\tilde{\theta}(T) \leq 0 \tag{28}$$

$$\|\tilde{\theta}\|_{\infty T} \leq \sqrt{\frac{-2(\beta_1 + \beta_2)}{\delta_m}} \tag{29}$$

Since β_1, β_2 and δ_m are constants, the inequality is true $\forall T \geq 0$, and then $\tilde{\theta} \in L_\infty$. Similarly, it is obtained that:

$$\|\tilde{\theta}\|_{2T} \leq \sqrt{\frac{-2(\beta_1 + \beta_2)}{\delta_{\Gamma K_0}}} \tag{30}$$

and consequently $\tilde{\theta} \in L_2$. The equation (10) can be put into form (31) , using (14).

$$\dot{\tilde{\theta}} = -\Gamma(Y^T s_1 + V^T W V \tilde{\theta}) \tag{31}$$

In [6], it is proved that if $Y(.)$ and $V(.)$ are *persistently exciting*, then $\dot{\tilde{\theta}} \in L_\infty$. Applying the Barbalat's lemma, since $\tilde{\theta} \in L_2 \cap L_\infty$ and $\dot{\tilde{\theta}} \in L_\infty$, it is concluded that $\lim_{t \to \infty} \tilde{\theta}(t) = 0$. ❖

3. SIMULATION RESULTS

In order to evaluate the theoretical results, the control law [1], [6], together the composite adaptive law (10), (15), was simulated on a two degrees of freedom robot structure presented in Fig. 2.

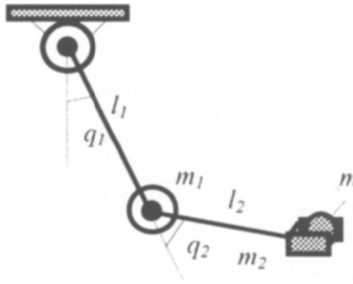


Figure 2. The two degrees of freedom robot structure.

The exact model parameters of the considered robot structure are presented in Table 1.

Table 1
The parameters of the robot model

Significance	Symbol	Value
Length of arms	$l_1 = l_2$	1[m]
Mass of joint 2	m_1	10[kg]
Mass of the effector	m_2	0.5[kg]
Manipulated payload	m_p	1.5[kg]
Coulomb parameter joint 1	$F_{c,1}$	5[Nm]
Coulomb parameter joint 2	$F_{c,2}$	1[Nm]
Viscous parameters	$F_{v,1} = F_{v,2}$	1[Nms]

The variable forgetting factor was generated with the following formula:

$$\alpha(t) = \alpha_0 \left(1 - \frac{\|\Gamma(t)\|}{k_0}\right), \ \alpha_0 \geq 0, \ k_0 > 0 \tag{32}$$

The friction torque in the dynamics model of the robot, (1), is detailed into its viscous and Coulomb friction components in (33).

$$F(\dot{q},\theta) = F_v\dot{q} + F_c sign(\dot{q}) \qquad (33)$$

F_v is the $(n\mathrm{x}1)$ vector of viscous parameters and F_c is the $(n\mathrm{x}1)$ vector of Coulomb parameters.

In the simulation experiment, it was considered that the payload mass and the parameters of friction torque cannot be exactly known but only estimated. Consequently, the vector of the estimated parameters was chosen as:

$$\hat{\theta}^T(t) = \begin{bmatrix} \hat{m}_p & \hat{F}_{v,1} & \hat{F}_{c,1} & \hat{F}_{v,2} & \hat{F}_{c,2} \end{bmatrix}.$$

For imposed sinusoidal trajectories of the robot joints:
$q_{d1}(t) = 0.34\,sin(0.3\pi\, t)$, $q_{d2}(t) = 0.62\,sin(0.7\pi\, t)$,
the values of the parameters were considered unknown and they were initialised by zero. The controller parameters were chosen as: $\Lambda = diag(5,5)[\mathrm{s}^{-1}]$, $K_d = diag(100,30)[\mathrm{Nms}]$, time constant of filter (11) $T{=}0.1$[s], $K_0 = diag(100,100,100,100,100)$, $\alpha_0 = 10$, $k_0 = 250$, $W{=}I$ (identity matrix). The adaptation gain matrix was initialised by: $\Gamma(0) = diag(3,\ 25,\ 45,\ 15,\ 15)$. The evolution of the position errors of the robot joints and the estimated parameters during 30s of simulation are presented in Figs. 3-6.

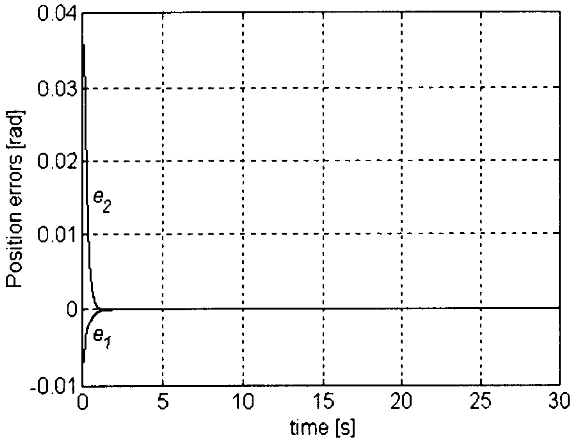


Figure 3. The position errors of joints 1 and 2.

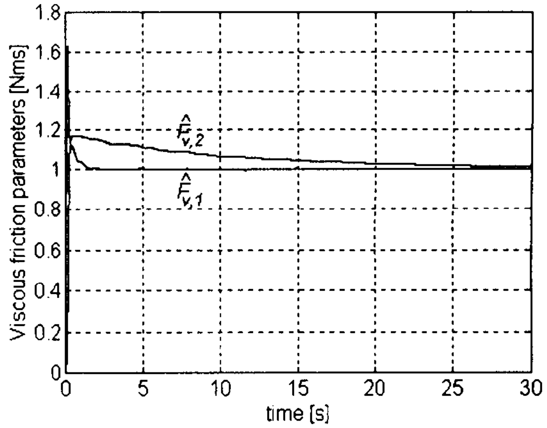


Figure 4. The estimated values of viscous friction parameters.

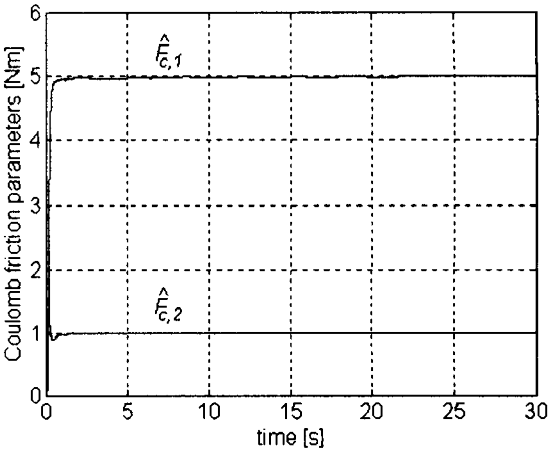


Figure 5. The estimated values of Coulomb friction parameters.

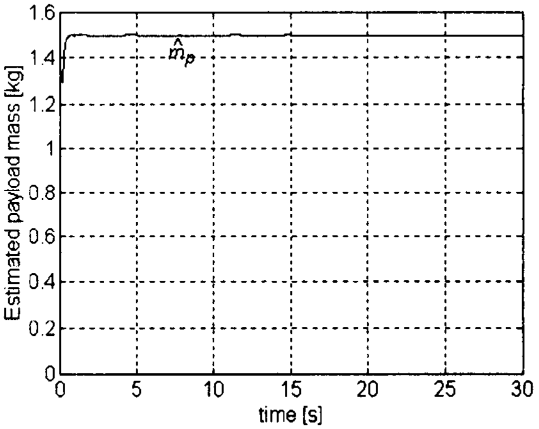


Figure 6. The time variation of the estimated payload mass.

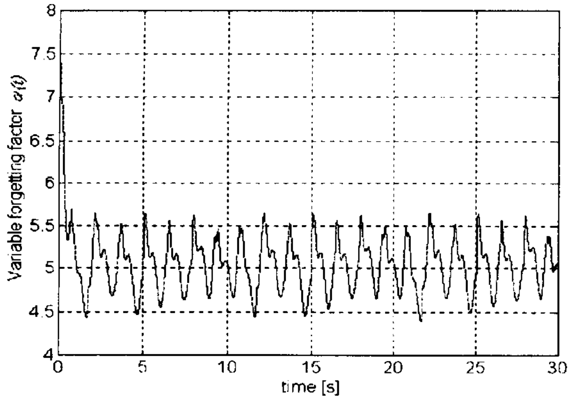


Figure 7. The evolution of variable forgetting factor.

The vector of the estimated values of the parameters at the end of simulation was:

$$\hat{\theta}^T = \begin{bmatrix} 1.49 & 1.00 & 4.99 & 1.00 & 0.99 \end{bmatrix},$$

compared to the actual values:

$$\theta^T = \begin{bmatrix} 1.5 & 1 & 5 & 1 & 1 \end{bmatrix}.$$

In Fig. 7, it is represented the evolution of the variable forgetting factor, using formula (32).

4. CONCLUSIONS

The motion performances of biological systems are a permanent challenge for the mechanical ones. In this paper, an adaptive controller, which takes into account the passivity property of the non-linear robot structure, is proposed. The stability analysis carried out for the closed loop is based on the passivity properties of the system substructures. Considering as parameters the payload mass and the friction parameters in the robot joints, the adaptive structure enables global asymptotic convergence to zero of the position errors and convergence of the estimated parameters to the exact ones. The simulation results conducted to the following remarks:

- The composite parameter adaptation law increased the speed of convergence and the precision of the estimated parameters, in comparison with gradient adaptation law, [8].
- The position errors are very small and comparable to those obtained in case of control law used in [1], [6] with an exact dynamic model of the robot structure.
- The oscillations of the variable forgetting factor are due to a relatively large initial value chosen for α_0. However, that enabled to observe the efficiency of the algorithm in maintaining a small position error together the precision of the estimated parameters.
- From the composite adaptation law, it can be obtained other variants of adaptation laws such as: $\alpha_0=0,\ W=0$ (the gradient adaptation law), $\alpha_0=0,\ W=I$ or an adaptation law with matrix Γ constant. Their performances can be evaluated comparatively.
- The method can be extended for the case when the parameters are slow varying in time.

Due to the high accuracy of the method in trajectory tracking, with uncertainties in the robot model, it can be suitable in robot industrial applications such as: laser cut or special technologies of welding. Another emerging field of application could be represented by medical experiments in surgery.

REFERENCES

1. Slotine, E J-J., W. Li, On the Adaptive Control of Robot Manipulators, Int. Journal of Robotics Research, 6, No. 3, (1987),pp.49-59.
2. Byrnes, I. C., A., Isidori, J.C. Willems. Passivity, Feedback Equivalence, and the Global Stabilization of Minimum Phase Nonlinear Systems, IEEE Transactions on Automatic Control, Vol. 36, No 11,(1991), pp. 1228-1240.
3. Ortega, R., M. W. Spong, Adaptive motion control of rigids robots: A tutorial, Automatica, Vol. 25, No.6, (1989), pp. 877-888.
4. Sadeh, N., R., Horowitz, Stability and Robustness Analysis of a Class of Adaptive Controllers for Robotic Manipulators, Int. J. of Robotics Research, Vol. 9, (1990), pp.74-94.
5. Desoer, C., A., M. Vidyasagar. Feedback Systems: Input-Output Properties, Academic Press, New York, 1975.
6. Slotine, E J-J., W. Li, Composite Adaptive Control of Robot Manipulators, Automatica, Vol.25, (1989), pp.509-519.
7. Bellman, R., Introduction to Matrix Analysis, McGraw-Hill Book Company, Inc., 1960.
8. Dumbravă, Şt., C., Pal, Robot Motion Control using Passivity Based Adaptive Controllers - Proceedings of the 3rd International Mechatronic Design and Modeling Workshop, Ankara, Turkey, September 15-18, (1997), pp. 137-146.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

New Workhorse for Lumberjacks

Pentti Vähä, Klaus Känsälä, Juha Kerva and Pekka Saavalainen

VTT Automation, Machine Automation
P.O.BOX 13023, Oulu, Finland

Abstract

Profitability in the first state thinning phase has remained low due to the small size of trees. Harvesters used for thinning today are generally intended more for heavy final felling than thinning. For this reason the efficiency at the thinning phase has stayed low resulting in rather high operating expenses. Therefore demands for forest machine being able to assist lumberjacks in their heavy thinning work have existed for some time.

1. NEW THINKING IN THINNING

After the new era started in Finnish forest industry in 1980's the main interest has been on paper and sawmill products. At the beginning the productivity in harvesting of trees increased very rapidly. However, since the first and second generation high tech harvesters have been brought to the market they are intended more for final felling than for thinning there has been no return to the past. These new machines have cut down the price to one quarter compared to that of manual workers. As a result jobs and lumberjacks disappeared from this branch. However, thinning work is very necessary to carry out in order to get good quality wood, and is almost impossible to perform with the existing machinery. These heavy and large machines are causing more damages than bringing any benefit when brought to young forests. Their productivity is also poor. Due to these reasons the first state thinning has not been carried out properly in large numbers of private and state owned forests. The overall quality of manual thinning has been very good compared to the other methods, but ever since the profitability has been the only criterion lumberjacks have almost disappeared from the forests. Apart from that, today's forest owners are more and more worried about environmental aspects like surface damages and wide driving tracks left at the thinning phase. This leaves market to small versatile thinning machines capable to operate on a little bit lower cost than manual workers, but still having an excellent working quality. So far no feasible partners for lumberjacks have been developed for making work more lightweight, efficient and profitable. The workhorse developed to perform work as a partner with the lumberjack accomplishes chopping, pruning and cutting to length as regular harvesters do. In addition to that this machine makes cutting onto its shoulder and transports the logs to the driveway and no forwarder is needed at the first thinning phase. Usually harvesters make cutting of trees onto the ground demanding a forwarder for transporting them to the collecting depot. This ability of its part reduces environmental damages and number of driving tracks. Furthermore this machine has very low surface pressure due to the total weight being only 2100 kg. Cutting thickness for chopping is 22 cm and maximum length for a log is 4.0 m. The harvester is able to carry 3 cubic meters of wood which means approximately one hour operating time before the load has to be emptied. For cutting operations automatic or semiautomatic operations in work boom control may be included to increase machine efficiency and easiness to use in addition to the usual manual control. For example, controlling only one joint can do cutting down a tree the other joint movements are synchronised according to this joint.

The user works with this harvester by exploiting a remote controller unit and selects his command place according to the task. The machine operates on both sides of the driving track, which may be at 60 meters distance from each other, Fig. 1. Usually they are situated 20 meters apart. If the distance to the nearest road is less than 200 meters it is useful and profitable to carry the logs with the harvester beside that road. In other cases the logs are piled along the driving tracks and are transported further on with a bigger forwarder.

2. ADVANCED PLATFORM FOR OFF-ROAD APPLICATIONS

One of the first tasks was to find a compact and small platform for the harvester in order to fulfil harvester specifications. This platform should be four wheels driven, have hydrostatic power transmission system with frame controlled steering [1-3]. Fortunately there existed a suitable commercially available vehicle. This vehicle was powered by a diesel engine employing hydrostatic power transmission. However, it was designed to operate only on roads and yards, not in forests. The manufacturer of this maintenance tractor, Oy Lai-Mu Ab, was willing to customise the product to be suitable for project purposes.

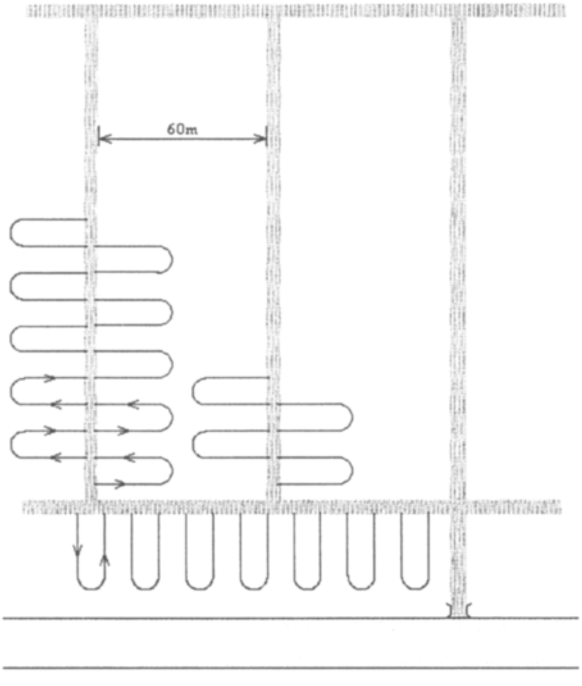


Fig. 1. Working with the harvester.

The cabin was removed and the manually operated hydraulics was replaced with an electrically controlled valve block. Also bigger wheels were installed in order to improve mobility in off-road driving circumstances.

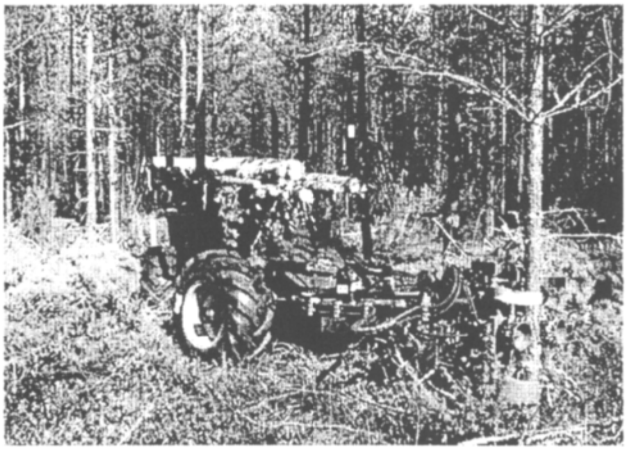


Fig. 2. The Harvester

A specially designed work boom with a harvester head is attached to the platform in order to do cutting, chopping and pruning in the same working cycle. Since the goal is to transfer the chopped logs out of the forest to the timber collecting depot the harvester has to be able to carry payload for some hundred meters. With this platform the average payload is estimated to be around 3 cubic meters of wood. In addition the machine is by estimation able to operate an hour time before the load has to be emptied. This makes it very competitive compared to other machines found on the market. It is also easily transferable from one place to another in a car driven trailer and has got good mobility on rough terrain. The most interesting applications are on tasks carriet out together with a human operator to provide energy or hydraulic power for jobs in hazardous environments. Stability is guaranteed by a flat body constraction placing all the mass between wheels giving a low center of gravity. The diesel engine as a balancing mass for the harvester boom is situated in the rear of the machine.

3. ANTI-SLIP SYSTEM

The platform is equipped with computer control and a special anti-slip control to improve the forwarding ability [4-5]. The target of the anti-slip system is to limit wheel slippage to the minimum in order to reduce surface damage and improve mobility in forest terrains. Slippage is a phenomenon which appears if the force applied to a wheel exceeds the surface friction. In vehicles with hydrostatic power transmission this causes an extreme power loss, since all the power produced by the hydraulic pump is dissipated by the slipping wheel causing it to accelerate strongly and to rotate with extremely high speed whereas all other wheels lose power and stop rotating.

One way to control slippage is to control individually the hydraulic pressure for each wheel. This method, however, relies on expensive hydraulic valves and is therefore not feasible on standard products due to its high costs. For this reason, the anti-slip control of the wheels is implemented by controlling individually existing brakes situated in each wheel. Implementation costs and system robustness had to be traded off in the design of the physical system [6]. Finally, only a cheap proximity sensor per wheel was used. The same unit 4serves as a velocity and acceleration sensor by counting the teeth of the gear mounted on the wheel shaft. The basic unit of the system, the axle controller, is based on Motorola 68332 processor. The same unit controls two wheels at the same time. The controllers are connected to each other via CAN-bus and they are constantly calculating vehicle speed, acceleration and turning angle and transferring this information to each other for control purposes in the global control mode. If slipping is detected wheel brake on slipping wheel is activated and slipping stopped by producing resisting force with the brake. The system is operating in the same manner than anti-slip systems presented by car manufacturers. The system illustrated in Fig. 3 is distributed and modular [7]. The number of axles can be adjusted from one to four. Thus, the maximum number of wheels in a vehicle can be eight.

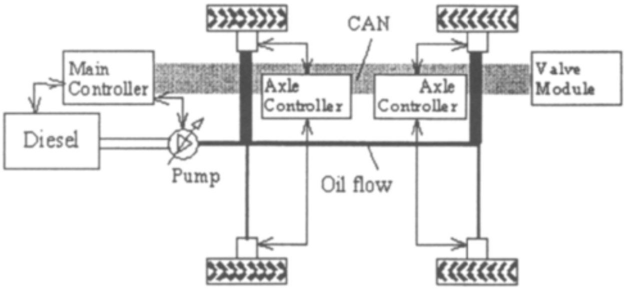


Fig. 3. The anti-Slip system

The CAN-bus protocol provides the physical communication layer. In order to be sure that the information is distributed a circular token passing schema has been implemented to allow collision and latency free deterministic exchange of information. The data to be exchanged includes mainly local speed and acceleration information for each wheel as well as brake settings and monitoring data. The token passing scheme uses token messages, which circulate through all wheel processors at a constant speed of 5-10 ms per cycle. This also guarantees detection of failing communication, and in case it is detected it activates the local control mode. In this mode anti-slip control is based on the information achieved only from one wheel.

4. SPECIALLY DESIGNED REMOTE CONTROL UNIT - SECURITY AND RELIABILITY

The machine does not have any cabin, but it is remote-controlled by the operator from the neighbourhood of the machine via a radio link. The command box hangs on the chest of the operator with shoulder straps. This gives the operator an opportunity to select a suitable command place according to the task. The link has been specially designed for outdoor applications in rugged environments. The reliability has to be extremely high, because the driver has to work at a short distance from the harvester. The reliability is based on three factors: constant monitoring of radio signal, special address function (fingerprint) for the transmitter and finally digitally decoded messages with CRC polynomials. The radio transmitter and receiver are standard components approved for use in EU-regions. The system is built using programmable FPGA-circuits. The transmitter reads 16-bit input port and decodes the signal, includes address bits, message frames and CRC-checksum bits. Then it activates the transmitter units and sends the coded message. The radio bandwidth is 400-480 MHz and the transmission power can vary from 4 mW to 500 mW. The receiver samples the radio signal and decodes the message. If the address bits do not match the message is rejected. If the signal is lost the receiver switches off after 0.5 second. The individual channels are updated 20 times per second. Up to 16 different addresses can be selected by the hardware.

The control of harvester is complicated and needs a large amount of analog and digital I/O. In our case there are 7 degrees of mobility (joints) to be controlled by using only two XY-joysticks. The solution was to group the controls so that the joysticks are controlling different motions depending on the state of the harvester whether it is stationary or moving ahead. In addition to this some movements had to be combined. After all this procedures the control could be realised using 5 analog channels (8-bit) and 20 digital I/Os. Four channels are reserved for joysticks and electric gas pedal is using the fifth one. There are several control modes for the user including automatic length measuring and stopping the log according to the pre set length, "floating" movements for the harvester head and automatic pulling down the tree after cutting it. In automatic mode boom joint movements are synchronized to each other. This means, that the operator needs only to control one joint of the boom while the other joints move automatically under computer control. Reference velocity and position values for these synchronized joints are calculated from the operator controlled joint values. Naturally all joints must have position transducers and be controlled by servo loops. Synchronized joint controls are used when approaching and gripping the tree before chopping it, and also when cutting into lenght and lifting wood to the shoulder. The latter task is combined together so that while tree is chopped or cut down, at the same time it controls the machine to rise it up to the position, where it can be cut to lenght by the harvester head to be fed to the shoulder. These kind of synchronized movements make operation of the boom moer easier and efficient to use.

5. FUTURE EXPECTATIONS

The thinning harvester developed is anticipated to penetrate into the thinning market and achieve its place as a partner for a group of lumberjacks by making their work easier, efficient and profitable. The size of the thinning harvester makes it easily transportable from one felling place to an other in a car driven trailer. Along with these it will increase the number of thinning made in Finnish forests, and thus result in wood with better quality. Embedded control systems from their part raise the value of the machine. Anti-slip control in employed to increase terrain mobility, and is very valuable in load transportation. Synchronised boom control makes the chopping and cutting operations easier and faster to perform, especially when cutting down and lifting wood onto shoulder.

6. CONCLUSIONS

A new mechatronic workhorse for assisting lumberjacks in thinning work was introduced. The main advantages of this system are its ability to move in young forests with minimum surface damages and openings for drive tracks. Embedded anti-slip control is used for increasing mobility and synchronised control from its part for making the chopping and cutting operations including the lifting of wood onto its shoulder faster and easier to perform. The workhorse introduced for thinning have been in field tests for some time and shown to be able to work as a partner for lumberjacks.

REFERENCES.

1 Clifford Mike, How to Select the Right Hydrostatic Transmission Circuit for Hydraulically Powered Vehicles, ASEA Inc. Houston Texas. 6 p. Off-Higway Vehicle Meeting and Exposition MECCA, Milvaukee 1979.

2 Friman Edward, Asplund Christer, Erikson Ulf, Mohr Erik. Evaluation of a Single Circuit All Hydrostatic Transmission with variable Wheel Motors, 27p., Sveriges Lantbruksuniversitet, Institutionen för skogsteknik, Uppsatser och Resultat nr 229, 1992 (in Swedish)

3 Friman Edward, Asplund Crister, Erikson Ulf, Mohr Erik. Desing and Test of a Controlled All Hydrostatic Transmission on a Forwarder. 34p., Sveriges Lantbruksuniversitet, Institutionen för skogsteknik, Uppsatser och Resultat nr 201, 1991 (in Swedish)

4 Friman Edward, Erikson Ulf, A Transmission with Varible Differential Locks and Automatic Slip Control. 13p., Sveriges Lantbruksuniversitet, Institutionen för skogsteknik, Uppsatser och Resultat nr 210, 1991 (in Swedish)

5 Hasemann, J.-M.; Känsälä, K. A Fuzzy Controller to prevent Wheel Slippage in Heavy Duty Off Road Vehicles. In Proceedings of the 44th IEEE Vehicular Technology Conference (VTC'94), Stockholm, Sweden, 1994, pp. 1108-1112.

6. Hasemann, J.-M.; Känsälä, K. An Embedded Fuzzy Anti Slippage System for Heavy Duty Off Road Vehicles, Information Sciences: An International Journal, Elsevier Science Publishing CO., New York, 1994.

7. Hasemann, J.-M.; Känsälä, K.; Pok, Y.-M. Analysis and Design of a Fuzzy Controller to prevent Wheel Slippage in Heavy Duty Off Road Vehicles. In Proceedings of the IEEE World Congress on Computational Intelligence/(FUZZ-IEEE 94), Orlando, Florida, 1994, pp. 1069-1074.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Intelligent Semi-Autonomous Vehicles in Materials Handling

Philip Moore[♣], Junsheng Pu[♣], Jan-Olof Lundgren[¤], and Sandor Ujvari[*]

[♣]Mechatronics Research Group, Faculty of Computing Sciences and Engineering,
De Montfort University
Leicester, LE1 9BH, UK

[¤]Volvo Automation, Engine Automation
S-541 87, Skövde, Sweden

[*]Centre for Intelligent Automation, Department of Engineering Science,
Högskolan Skövde
S-541 28 Skövde, Sweden

Abstract: An increase in functionality of Semi-Autonomous Vehicles (SAV) through the implementation of intelligent distributed control and smart sensing techniques is presented. In combination with a modular design approach, this facilitates system modification and improvement, combined with faster customisation of the platform. A distributed and reactive behavioural control architecture will be used to realise local autonomous navigation capabilities; improved operator interaction; self protection and safer operation. A virtual engineering environment based on a suitable computer-aided-graphics platform will be used for modelling the vehicle; the environment in which it can operate; and pre-emptive learning and training of responses / behaviours.

1.0 INTRODUCTION

The primary aim of this research is to improve the operation, safety, self protection, efficiency and flexibility of semi-autonomous guided vehicles / assembly carriers through intelligent control and smart sensing techniques in materials handling and related applications. This will lead to safer, highly effective and responsive man-in-the-loop[*] materials handling in manufacturing environments (e.g. automotive assembly) and other environments (e.g. baggage handling in airports).

The movement of materials within a production process has a very important function in manufacturing. One way to meet the needs for materials handling in industrial applications such as automotive manufacture and assembly is the use of assembly carriers and guided mobile robots. These are used to meet some of the materials handling needs, particularly in industries such as automotive manufacturing where the product and materials tend to be large and heavy which are difficult to manipulate by other means. Such handling systems need to be highly flexible to meet changing production schedules and transportation routings. The concept of intelligent semi-autonomous vehicles is both novel and timely for manufacturing and other sectors in meeting the need for improved performance, safer operation, flexibility and ease of use.

Older generations of mobile robots / vehicles (often referred to as AGVs) are typically wire guided with low functionality control systems. A significant step forward in development was the use of navigation sensors such as laser scanners in recognising beacons, or reflective tape set up in their operating environment, which allows the vehicle to follow virtual paths, such as NDC's free roaming system [1]. Many differing approaches have been adopted in recent years in the design of control systems for autonomous mobile robots, e. g. using horizontal decomposition of control systems into behaviours, as Brooks subsumption architecture, or more traditional vertical decomposition into functional modules, which are summarised in [2-5]. Much of the on-going research activity in the field can be found in 'Autonomous Mobile Systems, a study of current research' [6].

---

[*] Man-in-the-loop in this context covers a wide spectrum of application scenarios from the direct manipulation of the vehicle in a manual control mode through to autonomous behaviours of a vehicle in carrying out a defined task which are monitored from a remote location.

Incremental changes to both wire-guided and human-guided mobile vehicles can yield very significant improvements in operational flexibility, performance and safety, while more substantial modifications in the methods used to design and build the vehicles can open new markets in areas in which their use is presently limited or non-existent.

2.0 SYSTEM DESCRIPTION

In respect of its mechanical construction, electronics, control and software architecture, the complete vehicle is conceived on a modular, component-based design concept as illustrated in figure 1. The purpose of the modular approach is to facilitate ease of maintenance, system modification and improvement, combined with faster customisation of the finished product. The concept is illustrated by reference to the vehicle body, which consists of standard modules (or components) which can be arranged in different configurations to create a customised platform (e.g. which can vary in width, length, load capacity, height, drive configuration, etc.). Equally, the control and sensing systems can be configured, re-configured and extended to meet particular application requirements.

In general, the operating characteristics of semi-autonomous vehicles (SAV) are not particularly time-critical. Two servo control nodes are employed to perform motion control tasks in which close synchronisation is required, with dedicated microprocessors for PID-based motion control. The SAV must operate in a safe manner whilst operating in environments with personnel present. The man-machine interface (MMI) capabilities, which exploit the features of the integrated PC and distributed control network, represent a distinctive feature of the vehicle. Control nodes can be configured for particular interface functions (e.g. programmable switch panels, special displays, and so on), while intelligence and diagnostics can be incorporated in the individual control nodes and components (e.g. smart sensors with inherent validation capabilities). The PC-based supervisory tools can facilitate multiple levels of system access, enabling sophisticated diagnostics and maintenance support to be integrated. The designer is afforded complete freedom to design for optimum ergonomics, while multiple operating modes can be built into the vehicle and accessed in a variety of ways.

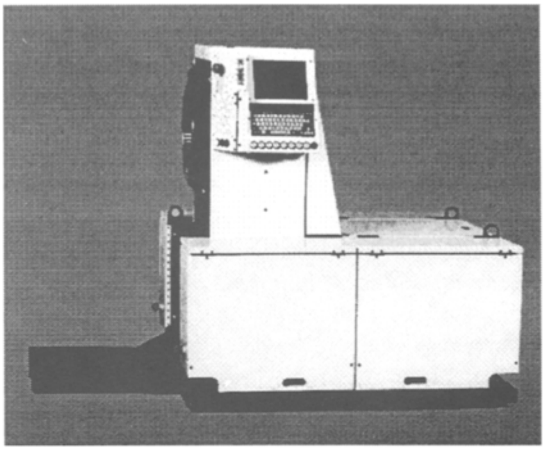


Figure 1. Assembly carrier vehicle. The load unit at the rear, to the left, can handle gearboxes, car-engines etc. On the side, the operator interface can be seen consisting of a PC screen / display panel and keyboard.

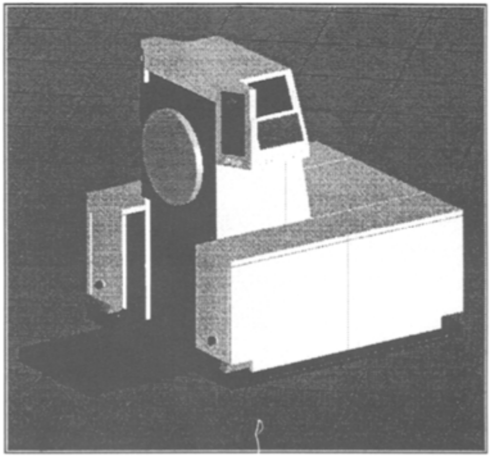


Figure 2. A simulation model of the vehicle.

The model is controlled with the same command language, provided in the form of a command file, as the real vehicle.

3.0 CONTROL ARCHITECTURE

The control system in this new vehicle concept is essentially distributed according to the physical location of the electrical components and control software. A range of control components based on the LONWORKS fieldbus are used on the vehicle platform. A standard industrial PC using the Windows NT operating system is used to perform supervisory, monitoring and display functions, see figure 3.

Some nodes are connected to sensors, such as wire-guidance antenna, laser range scanner, and nodes dedicated to man-machine interaction. The information is locally processed and provided via the network. The node that controls the drive and steering function of the vehicle also receives its commands via the network. The motion commands can come in automatic mode from the PC-node, or in manual mode from the operator. To co-ordinate the command execution, the PC-node works at a supervisory level.

The vehicle is programmed on the PC, using a script language in the form of a command file (see table1). The motion commands shows that the vehicle either follows a guidewire, or uses free navigation, through dead-reckoning. However, the script language structure simplifies the development of free-navigation functions using additional sensor information. An example is the laser range scanner [7] placed at the front of the vehicle at a height of 5 cm, which is an integral part of the safety system (see figures 4 and 5).

This architecture supports the fusion of sensor information, and arbitration of behaviours [5], e. g. the two commands follow wire and avoid obstacles.

4.0 DEVELOPMENT OF SMART SENSORS

The vehicle control architecture is designed to support the integration of a large range of sensors, e.g. for wire-guidance, laser range scanner, laser-scanner for detecting reflex tabs, ultrasonic sonar imaging systems, etc..

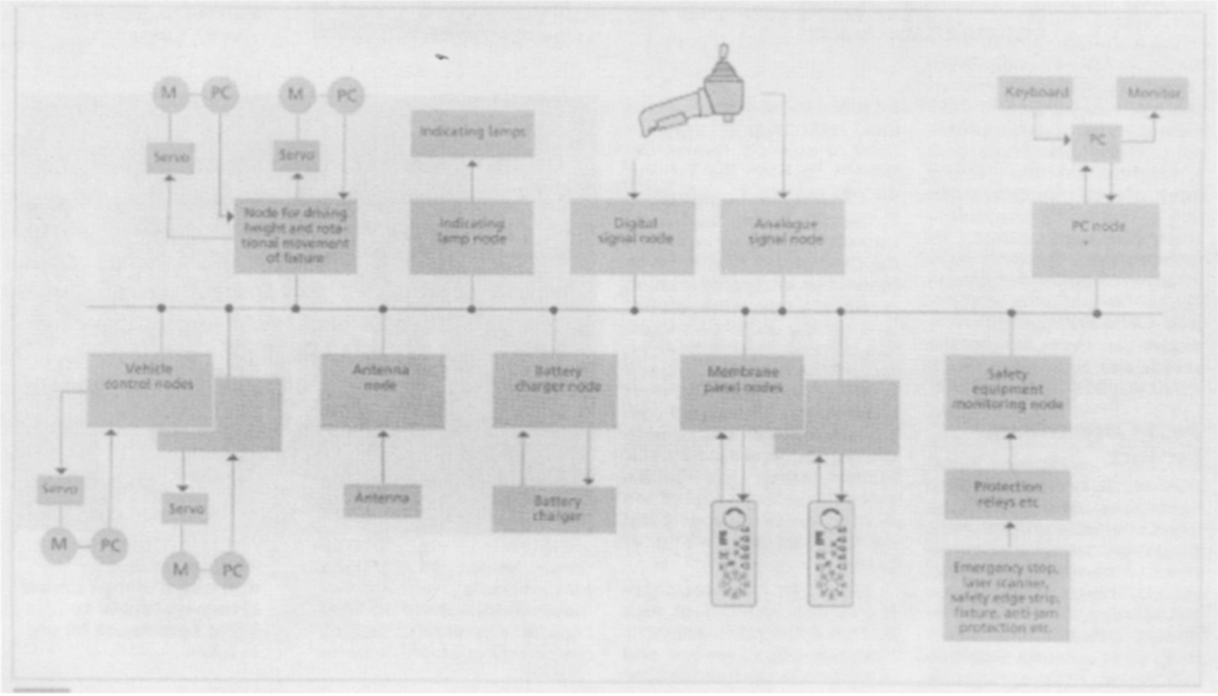

M
PC
Servo
M
PC
Servo
Node for driving height and rotational movement of fixture
Indicating lamps
Indicating lamp node
Digital signal node
Analogue signal node
Keyboard
Monitor
PC
PC node
Vehicle control nodes
Servo
M
PC
Servo
M
PC
Antenna node
Antenna
Battery charger node
Battery charger
Membrane panel nodes
Safety equipment monitoring node
Protection relays etc
Emergency stop, laser scanner, safety edge strip, fixture, anti jam protection etc.

Figure 3. The Lonwork-bus with dedicated processors forming distributed nodes of the control architecture. The vehicle control nodes are controlling the motion of the vehicle (steering and driving). During execution of a motion command, e. g. wire-following, the sensor information from the antenna is directly provided to the vehicle control node.

DriveWireToDistance [Percent of Max. Speed] [Distance]
DriveWireToPlate [Percent of Max. Speed] [Maximum Distance]
DriveFreeToDistance [Percent of Max. Speed] [Steering-Wheel Angle]

MoveFixture [Translation length] [Rotation angle]

Table 1. Example of the script language. The command is followed by one to three arguments, and controls the motion of the vehicle (line 1 to 3) and the load handler (line 4).

Different combinations and types of sensor for the SAV will be needed in different application environments.

A generic sensor model is being used to develop virtual models of the sensors for graphically based 3-D computer simulations. The simulation tool used is Cimstation Robotics [8]. In this way event based behaviours of the vehicle can be modelled in a virtual environment for purposes of evaluation and training. A detailed description can be found in [9].

The smart sensors used offer various levels of sophistication and use a mix of technologies such as wire-guidance antenna, laser range scanner, and ultrasonic sonar system. The wire-guidance sensor is one dimensional in nature (e.g. outputting one distance), the laser range scanner providing distance information in one plane is two dimensional, and the sonar array system provides three dimensional imaging.

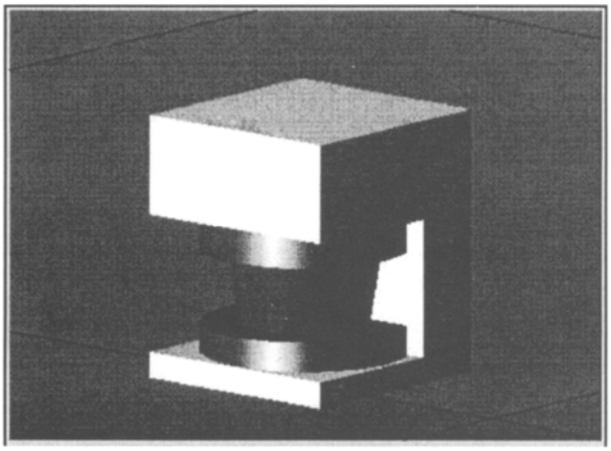


Figure 4. Sensor body of the laser range scanner

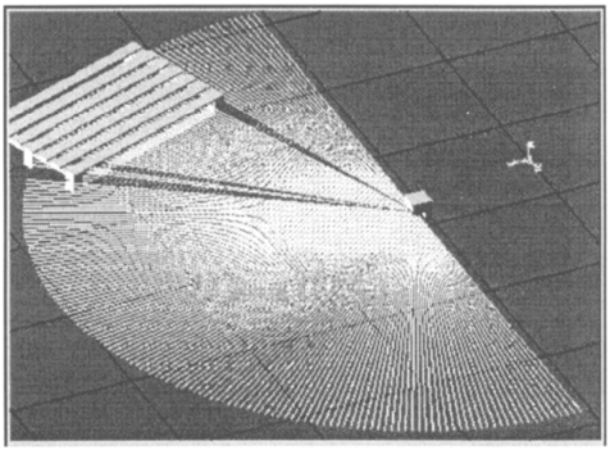


Figure 5. Laser range scanner model detecting a wooden pallet.

The first sensor to be modelled was a wire-guidance antenna, which detects the magnetic field generated around the guide-wire by alternating current typically in the kHz-range. The control loop typically consists of: an antenna attached to the steering wheel axis, and a PD-regulator that controls the wheel angle and speed of the drive-wheel. At some distance from the wire, the magnetic field becomes too weak to be detected and the sensor output goes to zero.

The laser range scanner measures distance in one plane through an angle of 180 degrees. The model simulates the operation by performing collision detection between 181 trace-lines (see figure 5) and surrounding obstacles. Since the distance readings provided by the laser range scanner are only just above floor level, additional information about obstacles in the path of the SAV are needed. An ultrasonic system is presently being developed that will provide 3D information about the area in front and to the sides of the vehicle.

To design mobile robots, such as semi-autonomous vehicles, a computer tool that is possible to use is computer aided robotics (CAR). CAR is originally used for the simulation and off-line programming of industrial robots. It can however serve as a platform for mobile robot simulation and emulation, where an application is developed, for the graphical event-based simulation of AGVs. The application will support the modelling of vehicles, sensors, control system and environment. Validity of results are tested using the real semi-autonomous vehicle. The purpose of the simulation is i) to develop 'pre-emptive learning'[10], to reduce training/learning of control algorithms required using real vehicles, ii) evaluate new vehicle configurations, e. g. sensor types and locations, iii) test the vehicle in new environments.

5.0 SUMMARY AND FURTHER WORK

The concept of an 'intelligent semi-autonomous vehicle' will be evaluated in terms of aspects such as ease and efficiency of use, the local navigation capabilities, the re-configurability of the system, etc.. Furthermore, in the virtual environment behaviours will be developed through 'pre-emptive' learning [10], evaluated and transferred to the vehicle. We envisage having a fully functional vehicle demonstrator to perform a full range of experimental studies.

The Intelligent semi-autonomous vehicles for materials handling project is designed to develop an integrated solution combining innovative control and sensing techniques (reactive behavioural control architectures and smart sensing systems) with state-of-the- art technologies (distributed control and virtual engineering design) in semi-autonomous vehicles. Volvo Automation's new, modular, semi-autonomous vehicles afford an ideal basis for this research project, the objectives of which are as follows:

i) To develop an intelligent control layer which can be implemented in the form of a reactive, behavioural control architecture suitable for 'semi-autonomous vehicles' which could include intelligent assembly carriers, intelligent AGV's which are capable of navigating locally, leaving the guide-wire to avoid obstacles in their path, and more sophisticated free-roaming vehicles with the aim of simplifying vehicle operation, improving self-protection, and affording safer operation and better functionality.

ii) To create a virtual environment in which the vehicle behaviour can be conditioned by 'pre-emptive learning', greatly reducing the need for on-site training and improving vehicle operation. Computer-aided graphical simulation tools will be required to model the vehicle, its sensors and environment, and their interaction in an event-based emulation closely resembling the real-life application.

iii) To design distributed components and system design tools for the development and integration of distributed control and sensing architectures for intelligent semi-autonomous vehicles.

iv) To develop smart sensors and sensing techniques suitable for intelligent vehicle guidance systems (novel digital signal processing techniques for sonar array sensors will be the subject of particular study).

ACKNOWLEDGEMENTS

The research programme described in this paper is funded by the INCO-Copernicus programme IC15CT960726 and The Swedish National Board for Industrial and Technical Development, NUTEK. The project also involves major contributions from Volvo Automation (Sweden), Kaunas University of Technology (Lithuania), TCC and MSTU Bauman (Russia).

REFERENCES

[1] NDC Co. AB, SE-429 80 SÄRÖ, SWEDEN, http://www.ndc.se/

[2] Brooks, R. A., *A Robust Layered Control System for a Mobile Robot,* IEEE Journal of Robotics and Automation, Vol. RA-2, No. 1, pp 14-23, March 1986

[3] Corfield, S.J., Fraser, R.J., and Harris, C.J., *Architectures for Real Time Intelligent Control of Autonomous Vehicles*, Computing and Control Engineering Journal, Vol.2, No.6, pp 254-262, November 1991

[4] Arkin, R. C., Murphy, R. R., *Autonomous navigation in a manufacturing Environment*, IEEE Transactions on Robotics and Automation, Vol. 6, No. 4, pp 445-454, August 1990

[5] Saffiotti, A., *The Uses of Fuzzy Logic in Autonomous Robot Navigation: a catalogue raisonné*, Soft Computing, Vol. 1, No. 4, Springer Verlag, 1997

[6] Uhlin, T, and Johansson, K., *Autonomous Mobile Systems, a study of current research*, Technical Report, NUTEK, R 1996:4

[7] SICK AG, Nimburger Straße 11, D-79276 Reute, http://www.sick.de/english/

[8] Cimstation Robotics, Silma Inc., 150 Rose Orchard Way, San Jose,California 95134, USA http://www.silma.com/

[9] Sandor Ujvari, Patric Eriksson, Philip Moore and Junsheng Pu, *Simulation and Emulation of Sensor Systems for Intelligent Vehicles*, Mechatronics 98, Sept. 9-11, 1998, Skövde, Sweden, Elsevier

[10] Moore P R, and Eriksson, P., *A Role for 'Sensor Simulation' and 'Pre-emptive Learning' in Computer Aided Robotics*, Proceedings of 26th International Symposium on Industrial Robots (ISIR'95), Singapore, 4-6 October, 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Remote radiation mapping and preliminary intervention using collaborating (European and Russian) mobile robots

Leon Piotrowski[a], Elgan Loane[b], Marc Halbach[c] and Nikolai Sidorkin[d]

[a]Electricité de France, Direction Etudes & Recherches, 6 quai Watier, 78400 Chatou, France.

[b]Kentree Ltd., Kilbrittain, Co. Cork, Ireland.

[c]Université Libre de Bruxelles, P.O. Box 106, 50 avenue F. D. Roosevelt, 1050 Brussels, Belgium.

[d]NIKIMT, 43 Altufievskoe Chaussee, 127410 Moscow, Russia.

Potential applications of mechatronics in mobile robotics are illustrated by a description and analysis of a mission that took place in a nuclear power plant.

INTRODUCTION

Mechatronics has numerous potential applications in mobile robotics for the nuclear industry. This is illustrated by a concrete example in which two collaborating robots were recently deployed within the Russian nuclear plant Smolensk as part of the end-of-project demonstration for the European project TELEMAN-IMPACT.

The robots developed by the IMPACT project were simple manually-teleoperated vehicles. Their use, however, in a realistic hostile environment gives valuable insight into which robotic features need improvement in order to render future missions easier for the operator.

The conclusions cited in this paper are typical end-user comments about the performance of intervention robots within hostile nuclear environments.

Part I of the manuscript briefly presents the IMPACT mobile robots and the mission that confronted them at Smolensk. Part II represents a "post-intervention analysis", from an end-user's viewpoint, and lists some robotic features that could benefit from mechatronics-type solutions.

Comments are also made on the importance of devising some sort of *strategy* whereby two or more simpler robots can work together. When applied to missions involving remote teleoperation within hostile environments, such an approach may well be preferable to one that over-equipes a single machine with elaborate technologies.

PART I
THE ROBOTS AND THE MISSION

1.1. The IMPACT project

The primary objective of the IMPACT project was to develop a light-weight and inexpensive mobile robot (demonstrator) that can be used for rapid inspection missions within nuclear power plants. These interventions are to cover normal, incident and accident situations and aim at primary reconnaissance or data-collecting missions.

The main benefits to be obtained from using the IMPACT robot are (i) plant personnel would have rapid access to information describing the state of a suspected incident, and (ii) the security of the plant would be improved by more frequent surveillance missions while, at the same time, avoiding human radiation exposure.

1.2. The Smolensk mission

The theme of the Smolensk demonstration was preliminary inspection and intervention by teleoperated mobile robots following an incident on a contaminated circuit.

Test site : Smolensk nuclear plant, Unit 2, turbine hall near turbine #4. The test site was in the radioactive zone of the plant.

The command post was located in a sheltered area behind a thick concrete wall (no radiation exposure for the operators) and the intervention was performed remotely.

The distance between the operators and the actual intervention site was approximately 12 m.

Demonstration : The IMPACT demonstration was composed of 2 independent but consecutive missions :

Mission #1 : Remote radiation mapping and radioactive source ("hot spot") localisation in 3D by the IMPACT robot equipped with a directional radiation sensor.

Mission #2 : Construction of a protective lead shield around one of the identified "hot spots" and verification that the ambiant radiation level has been reduced.

The second mission was executed remotely by 2 mobile robots working in collaboration : a Russian intervention robot equipped with a manipulator arm and carrying lead bricks and the (European) IMPACT robot of Mission #1 above.

During the second mission the IMPACT robot functioned as a "side-observer" and (i) supplied the (Russian) operator with an external view of the intervention robot (thus aiding in the building of the protective shield), and (ii) was available for any additional radiation measurements to confirm that the lead shield was effective in reducing the ambiant radiation level of the test site.

1.3. The IMPACT mobile robot

The IMPACT mobile robot is an extension of a light tracked-platform that has been developed by Kentree (Ireland) for civil security and military applications. The project improved many features of this robot, thus making it more suited to inspection missions within nuclear plants.

Special attention was given to producing a flexible but robust platform which could be deployed in different ways depending on the plant operator's needs and/or on the constraints of a given mission :

A. Mechanical Capabilities

The IMPACT robot has a 3-segmented, 6-caterpillar-tracked body with articulations between the segments. Other characteristics are :

dimensions	:	135 (L) x 41 (l) x 25 (H) cm
weight	:	70 Kg (approx.)
speed	:	12 cm/sec. (approx.)
carrying capacity	:	80 Kg
battery autonomy	:	2 hours (approximately).

Battery autonomy depends on the additional sensors being carried and on the nature of the obstacles that have to be overcome.

Key features of the mobile platform are its low center of gravity and its 3-segmented, articulated body. These endow it with both agility and stability; properties that the project considered essential for any primary reconnaissance robot for nuclear plants.

B. Ability to Divide Itself into Two Robots

During a given mission, the operator is able to detach *remotely* the rear segment of the platform and thus obtain 2 separate mobile robots. The detached mono-segment can serve several functions: to offer the operator an external view of the main (2-segmented) module thereby facilitating its teleoperation, to act as a possible rescue for the main module (e.g. to recharge the batteries of the main section, if the mono-segment is connected to the command post by a cable), to allow the robot to pass through tight passages or sharp corners and to supply the operator with multiple views of the environment.

The detached mono-segment was also designed to function as a radio communications relay for the independant (2-segmented) module.

The operator can also remotely re-attach the detached segment to the main module and thus regain the original 3-segmented configuration.

C. Multiple Teleoperation Cameras

Ease of teleoperation and pilot comfort are assured by 6 CCD cameras placed at various locations around the 3-segmented platform (front, rear, right & left for the 2-segment module and front & rear for the 3rd segment). The main front and rear cameras supply colour images, the others being miniature B/W "navigation" cameras to help driving in cluttered environments or to help in disconnecting/reconnecting the 3rd segment from/to the main module.

A "quad-split" video transmission system is used to display simultaneously multiple (4) images at the operator post.

D. Cable or Cable-Free Operation

The 3rd segment may be connected by a cable to the command post. Alternatively, radio communications can be used between the robot and the operator. Radio transmission is always used between the main module and the 3rd segment. Communications consist of teleoperation commands, video transmission and data transmission.

E. Teleoperation Post

The command post is comprised of a console of switches (manual teleoperation) developed by Kentree and a PC-based display screen (software) developed by Université Libre de Bruxelles. The display screen offers the operator a combination of video and graphic information. The latter shows the status of the IMPACT robot (e.g. inclinations of the platform's 1st and 3rd segments, platform pitch and roll angles, battery level, etc.).

F. Gamma-radiation sensor

The radiation sensor, developed by NIKIMT (Russia), is an integrated unit (28 Kg) capable of scanning a target environment in order to localize radioactive sources in 3D space.

The sensor had a 10° cone field-of-view and the integration time for a point radiation measurement could be set by the operator (10 seconds used at Smolensk).

In addition, the operator was able to orient the gamma sensor remotely from the control PC to any desired pan (range : -160° to +160°) and/or tilt (range : -25° to +90°) angle.

The gamma sensor was carried by the IMPACT robot (which was cable-free at Smolensk) but had to be connected to its command PC by a RS232 cable.

1.4. The intervention robot

The Russian intervention robot consisted of a single main platform, manipulator arm, 2 B/W CCD cameras, lighting system and provision for carrying approximately a dozen lead bricks. It is capable of communicating with the operator via cable or radio. Radio communications was used at Smolensk.

The general characteristics of this robot are :

overall dimensions	:	140 x 70 x 60 cm
weight	:	250 Kg
speed	:	0.05 to 1.5 m/sec.
carrying capacity	:	150 Kg (platform); 20 Kg (arm)
duration of work	:	6 hours
manipulator arm	:	axis-by-axis teleoperation.

1.5. Execution of the Smolensk mission

The total time required to perform the radiation mapping operation was approximately 1.5 hours. No trial runs, or "practice sessions", were performed at Smolensk.

The protective lead shield (built from lead bricks) reduced the ambiant radiation level by a factor of 270. This phase of the mission lasted about 2 hours. All manipulations were performed remotely. Again, no on-site practice sessions were performed prior to this intervention.

At no time during the execution of the Smolensk mission did the operators have a direct view of the robots.

PART II

POSSIBLE MECHATRONICS APPLICATIONS

An end-user's view

The following remarks suggest (robotics) areas that might benefit from the application of mechatronics-type technologies. They are taken directly from the Smolensk demonstration.

2.1 Improvements to a single mobile robot

The 3-segmented articulated body of the IMPACT robot proved efficient for overcoming a variety of obstacles. The weak point of this feature is that the operator must interrupt his driving and manually raise/lower the front and rear segments. Less stress and greater operator comfort would be achieved by teleoperation aids that automatically raise the front segment when confronted with an obstacle or lower a segment if it is not "touching ground". Experience from other missions has also shown that automatic stair-climbing and door-opening capabilities would be much appreciated.

It is well-known that long missions (e.g. 100 m or more) require an umbilical cable. A "cable-handling mechanism" would free the operator from constantly checking that the trailing umbilical was not snagged in the cluttered environment. This was of some concern at Smolensk since the radiation sensor had a cable attached to it.

The IMPACT robot can remotely detach its third segment and thus become two independent mobile vehicles. The inverse is also possible. Both operations are manually controlled from the operator post and require a certain degree of training. The re-attaching procedure, in particulier, is somewhat delicate because it requires the alignment of the two modules. Automating this action would be of considerable help to the operator.

Not all driving is performed in cluttered environments. Situations do occur when the mobile vehicle is driven in relatively open spaces. Automatic aids for "continue in a straight line" or "follow this wall" are examples of simple functions that would help the operator.

Good vision is essential for driving or inspecting. If radio communications is used then there is a serious risk of interference from the surrounding (concrete and metallic) structures. The result is often a momentary loss of imagery which, at the best, is annoying. If however the losses in communication are frequent, or for relatively long times, then driving becomes impaired. A system that suppresses the display of scrambled images and (temporarily) replaces them by the last good quality image (even though in a frame frozen format) would help the operator. This solution would be feasible for a few seconds because indoor mobile robots are not driven at high speeds.

If a serious communications loss occurs then there is a risk of loosing the vehicle. An "automatic reverse action" that retraces the last 1-2 m could re-establish communications and save the robot.

The intervention robot at Smolensk had the task of building a protective wall from lead bricks. This teleoperation action was completely manual. Aids in grasping bricks and posing them on top of each other would render this task easier. Much time and operator energy was consumed in putting up a simple protection around the identified radioactive source.

2.2 Collaborating robots

The Smolensk mission was greatly simplified by using simple dedicated robots working in collaboration.

This approach of task sharing between relatively simple machines is particularly relevant to remote teleoperation within hostile environments and offers numerous practical advantages.

For example, using more than one robot implies that the operator may have multiple viewpoints of certain manipulative tasks.

More importantly, it avoids the need to build complex robots that would be expected to execute a number of different functions. Task sharing between cooperating specialised robots hence simplifies the mission and increases the probability of success.

Finally, operator stress is reduced by simplifying the machines that are deployed.

A distinction must be made between equiping robots with advanced *technologies* and the *strategy* that machines could (should) use to solve complex problems. The Smolensk demonstration showed that a single vehicle, no matter how sophisticated, should not be expected to overcome all difficulties on its own.

In some cases, mechatronics should therefore also include machine interaction and coordination.

ACKNOWLEDGMENTS

This work was partly sponsored by the European Commission's (DG XII) research programme on robotics for the nuclear industry (TELEMAN). Additional support also came from the PECO programme (Scientific Cooperations with Central and Eastern European Countries).

Special thanks is also given to the personnel of the Smolensk plant for their kind cooperation.

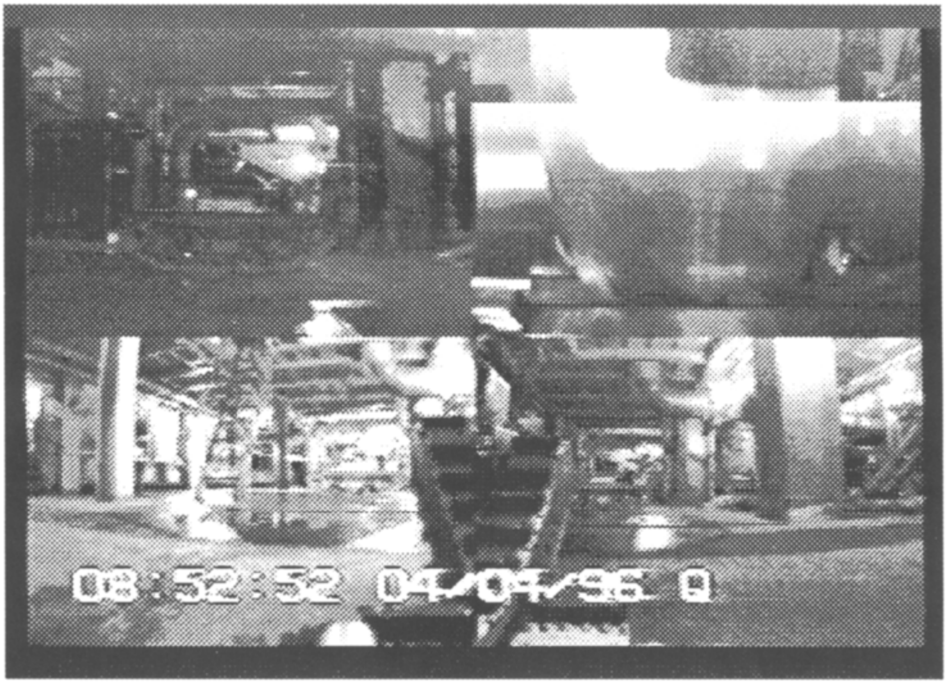


Photo 1.
Quad-split video display showing the simultaneous presentation of 4 views captured by 4 of the IMPACT robot's CCD cameras. The photos were taken when the robot entered the intervention site.

Top left image (front camera view) : The Russian intervention robot is on standby.
Top right image (rear camera view) : Part of the 3rd segment [The rear view from the 3rd segment is not displayed here].
Bottom left image (left-side camera view) : Left forward-looking view of the intervention site.
Bottom right image (right-side camera view) : Right forward-looking view of the intervention site.

The side images are a valuable aid for remote teleoperation because they cover a wide field-of-view.

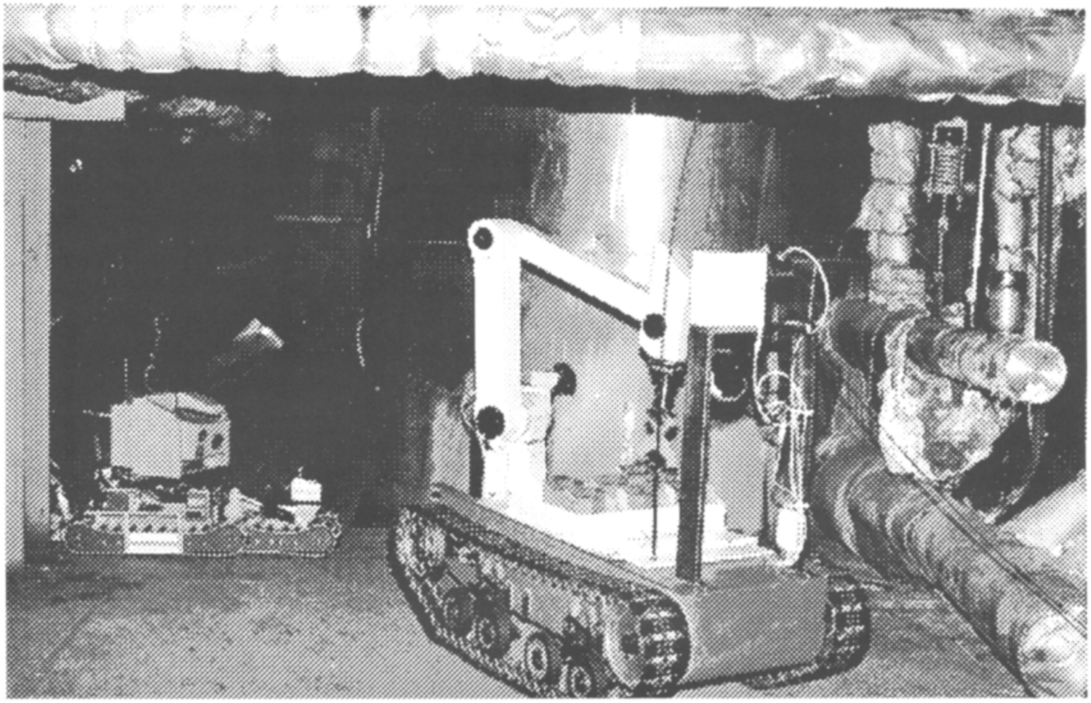


Photo 2.
Having completed a radiation survey and localised the radioactive sources, the IMPACT robot (left) assists the intervention robot (right) in placing lead bricks around a source in order to reduce the ambiant radiation level. The IMPACT robot assists the Russian operator by supplying him with an additional side view (see photos 3 and 4). The IMPACT robot also continues to perform radiation measurements to verify that the protective lead shield is effective.

RIGHT
11:45:09 04/04/96 0

Photo 3.

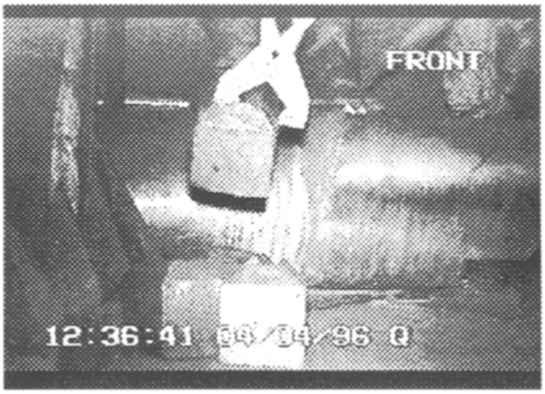


Photo 4.

The above images were taken by the IMPACT robot behaving as a side observer to the intervention robot. These views were in addition to those supplied by the intervention robot itself and represented a valuable operator-aid for remotely placing the lead bricks around the radioactive source.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Development of mobile robots for mines detection

E. Colon[a], Y. Baudoin[a], P.Alexandre[b]

[a]Department of Applied Mechanics, Royal Military Academy,
Avenue de la Renaissance 30, 1000 Brussels, Belgium

[b]Department of Mechanical Engineering and Robotics, Université Libre de Bruxelles,
Avenue F.D. Roosevelt 50, 1050 Brussels, Belgium

Antipersonnel mines kill or mutilate tens of people every day. Humanitarian deminers use the classical manual methods because heavy demining vehicles cannot achieve a satisfying destruction percentage. This work is very slow, tedious, dangerous and costly and furthermore, the detection is not always reliable. Improvements can be made by developing new sensors and by automating the detection sequence. The Robotics workgroup of the Belgian Hudem project focuses on the development of low cost vehicles that could be equiped with mine detectors. In this paper, we shortly review the main characteristics of such platforms. We then show that the architecture of the vehicle is also determined by the chosen detection strategy. Afterwards, we describe the vehicles we are modifying or building. All kinds of propulsion are considered: wheels, tracks and legs.

1. INTRODUCTION

More than 125 million of unrecovered anti-personnel (APM) and anti-tank (ATM) mines can be found in more than 60 countries. In countries (like Cambodia, Afghanistan, Kurdistan etc.) where the presence of landmines became a part of the everyday life, the consequences of the landmines problem on humanitarian and environmental level are enormous. First of all, the physical damage that very often can't be followed by a proper medical treatment, and secondly psychological damage. The other problem is the economic factor. Landmines can also cause underproduction, a further barrier to development.

We can distinguish the military and the civilian mine clearance. The military demining operations accept low rates of Clearance Efficiency (CE). For these purposes it is often sufficient to punch a path through a mine field. For the humanitarian demining purposes, on the contrary, a high CE is required (a CE of 99.6% is required by UN). This can only be achieved when using the hand clearance method.

A manual mine clearance campaign consists of:

1. Mine field marking
2. Detection of tripwires
3. Mine detection
4. Mine prodding (Figure 1)
5. Destruction of mines.

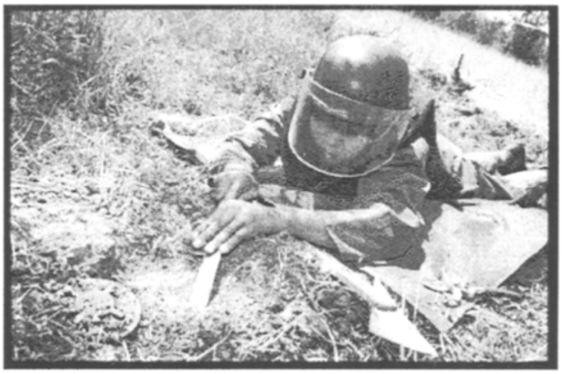


Figure 1. Manual prodding

stages 2 to 5 are repeated for corridors of 1m wide by steps of 0.5m. A team of 30 well trained people is able to clear approximately 1000m2 a day. This method has not changed since the World War II.

This method of demining is very expensive. It costs between 3 US$ and 25US$ to lay one mine, but up to 1,000 US$ to clear one. The main reason for this difference is a big false alarm rate because of the metal detector that is commonly used in phase 3. (from 500 to 1000 false alarms for 1 true).

A lot of new sensors or techniques are being experimented and today, the most promising technique seems to be the combination of a metal detector with a GPR (with possibly a third sensor: IR or magnetometer).

Our project, named HU(manitarian) DEM(ining), is composed of 4 groups:

- Radar workgroup,
- Signal Processing workgroup,
- Robotics workgroup,
- Support workgroup.

The Robotics Work Group has received the following tasks:

1. a study of the sustainable mechanical solutions for the humanitarian mine clearance,
2. the development of modular low-cost mobile platforms,
3. the development of control algorithms using images of the environment and data coming from mine sensors.

After this relatively long introduction, we present the requirements a mine detection vehicle should meet. We then show how the detection scenario influences the characteristics of the vehicle. Afterwards, we describe the wheeled robot and we say some words on our legged and tracked prototypes. Before concluding, we make some considerations on the Man-Machine Interface.

2. REQUIREMENTS

The study started with a state-of-the-art survey [1-3], many discussions with professional deminers and some visits to countries concerned with this problem (Cambodia, Croatia).

It emerged from the analysis carried out that a mechanical system for mine-detection must have the following qualities:

- reliability, both mechanically and in the detection of mines
- high resistance to explosions
- ease of implementation
- good range (running at least half a day without refueling)
- easily transportable by light utility vehicle
- rapidly repairable in the case of breakdown or partial destruction
- low cost
- resistance to extreme climatic conditions (heat, damp, sand etc)

It seems totally utopian to develop a vehicle that could be used in all circumstances: desert, wood, bush, hot or cold, wet or dry weather. Furthermore, the characteristics of the vehicle will also depend on the way we will use it as explained in the next paragraph.

3. DETECTION STRATEGY

We have chosen to follow the existing demining procedures. When a mine has been discovered, the position is indicated with a beacon and the operation goes on in another corridor. Mines are destroyed at the end of the day (or half day). Different scenarios can be considered when replacing the man by the machine.

In the first scenario we accept to sacrifice the robot; in this case we take the risk to roll or walk over a mine and the vehicle must be disposable. In this case the sensors can be put everywhere on the robot.

The second scenario tries to preserve the robot. In this case, we cannot simply replace the man by a robot, because we will have to stop the robot each time something has been detected, to go backwards and to maneuver to start in a new area. We suppose also that the mine zone is bordered by an area free of mines (it is always the case). The robot will follow the mine field while scanning the ground laterally. Doing so, we don't have to stop the robot each time something has been detected.

4. DESCRIPTION OF THE ROBOTS

Each kind of motion transmission, ea. wheels, tracks and legs, has its advantages. In order to evaluate which solution is the most adapted to a given environment, we are working on different prototypes.

4.1. Tridem

4.1.1 Design

Tridem is totally new wheeled machine. In order to obtain a large mobility, we have decided to equip the vehicle with three independent drive/steer wheels. The wheels are connected by a triangular frame. Each wheel has 2 electrical motors. The frame will support the control electronics and the batteries. The control system is a microcontroller

M68332. The vehicle can communicate with a remote computer through a serial link.

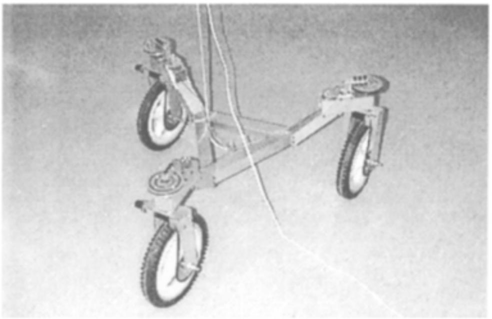


Figure 2. The Tridem

The robot has been designed to have a 20 kg payload and a speed of 0.1 m/sec. Aluminum sections have been used for building the frame of the robot; it give our robot a high rigidity and a low weight. The wheels diameter is 12 inches. It should be enough to move on rough terrenes with holes and bumps of maximum 10 cm.

The electrical motors are manufactured by MAXON. The driving motors are equipped with a tachometer and the steering motors with a digital encoder. Both are directly connected with planetary reduction gears. The characteristics of the motors are listed in table 1.

Table 1
Characteristics of the motors

	Driving motor	Steering motor
Power	15W	6W
Stall torque	87 mNm	43 mNm
Reduction ration	246:1	33.2:1

Chains and pinions are used between the driving gears and the wheels. The steering planetary gear and the rotation axes are linked by low backlash chain/pinion systems. As the rotation axis is aligned with the contact point of the wheel with the ground, the wheel can turn in place without effort when the vehicle is still.

The power is transmitted to the driving motors with copper brushes. This gives the wheels a complete rotation freedom. An optical sensor gives an absolute reference for the wheel positioning.

A TT8 microcontroller card is used for motion control. This card is fitted with a Motorola 68332 microcontroller, a 8 channels 12-bit AD converter and 256 K of RAM. It is a low consumption device, with a user-controlled running frequency. The multiple Time Programmable Unit are very convenient for controlling electrical motors. They can be used as digital input/ output, for signals monitoring or PWM.

The wheels can be removed and replaced very easily because of the modular conception. All wheels are identical, they are fastened to the frame with fast screw connections. The wires (signals and power transmission) are connected with standard connectors (DB15).

4.1.2 Remote control and programming

The chosen architecture allows a high mobility. The free positioning of the wheels gives a large number of possible configurations as illustrated by the figure 3.

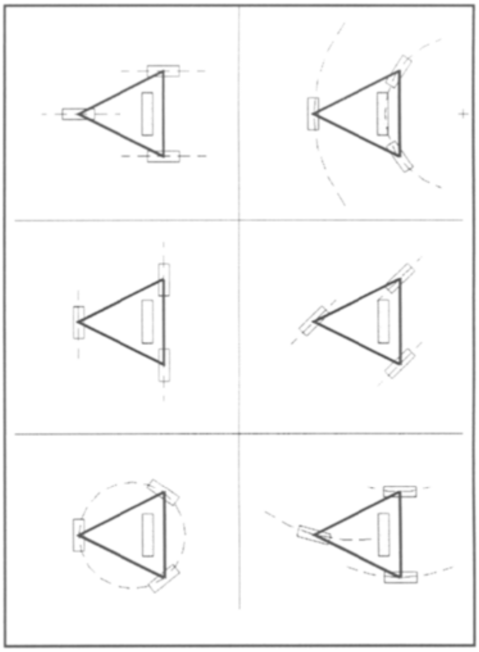


Figure 3. Some possible motions

Programs have been written for simulating, testing and controlling the robot. The figure 4 shows the user interface for the "single action" version. With this program we can control the wheels individually.

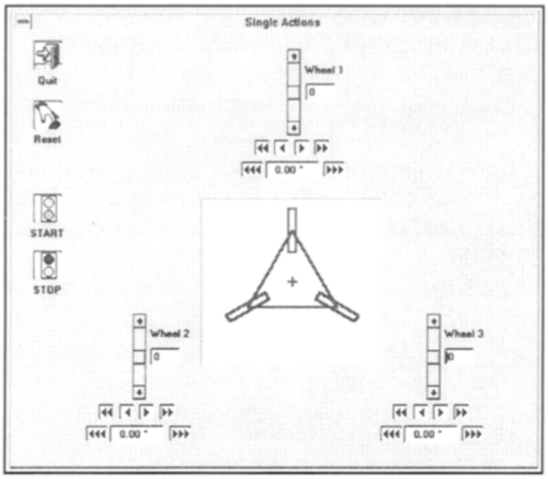


Figure 4. Single action program

The second program computes the wheels' speed and orientation for a given path. You can choose the global speed, the main orientation for a straight motion, the rotation speed and the radius for a circular motion. These two motions can be combine with a rotation of the vehicle about its geometry center.

The user can now control the robot with a self-made joystick. The x-y axis control the direction of the vehicle. The stick has a supplementary degree of freedom (a rotation about its axis) to change the configuration of the wheels.

Indoor and outdoor trials have been successfully completed. Further tests will be done on uneven terrains.

4.2. Walking Robots

In very unstructured environments and if the security of the human deminer has to be improved, the use of a light wire-guided (avoiding navigation problems) legged robot can be considered.

We are working on the modification of a six-legged and a four-legged robot.

The aim is to control the movement of the robot at the same time as stabilizing its seating: the legs will be equipped with inclinometers connected to a Motorola 68332 microcontroller: in order to limit the costs, simple on-off valves have been used to set the legs in motion or stop them. The sensor will be placed at the extremity of a boom elastically bound with the structure of the robot and also with the ground via a rolling spherical joint. A particular attention has also been given to the design of flexible feet and the limitation of the pressure applied on the ground, that must remain lower than 5 bar.

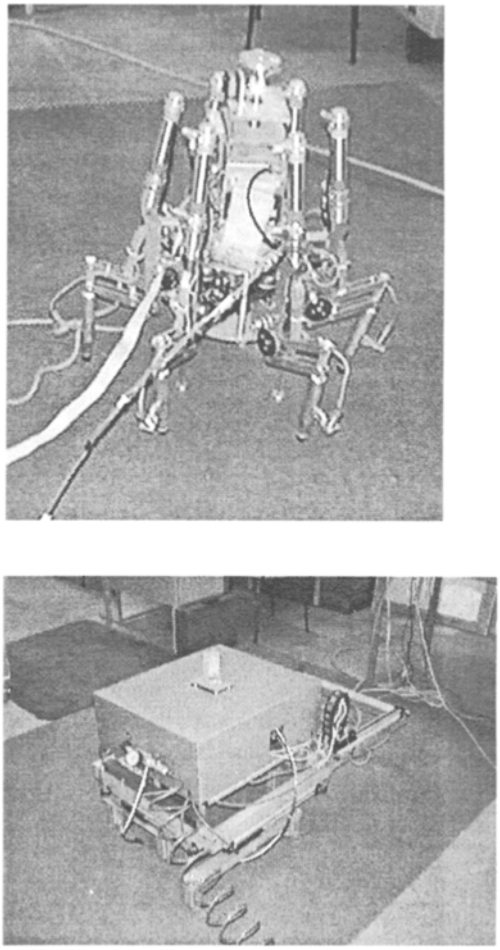
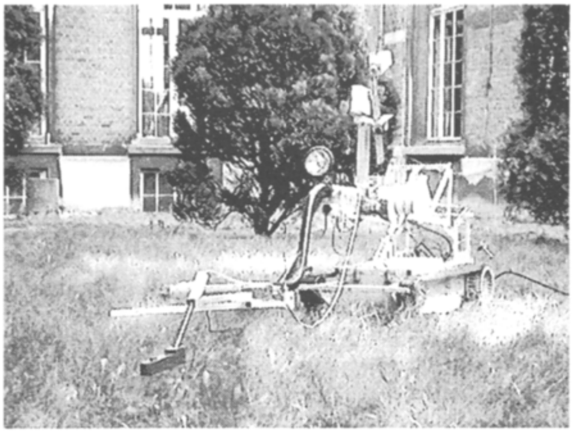


Figure 5. The 6 and 4 legs robots

4.3 Tracked Robot

The advantage of tracks is that they can propel the vehicle on all types of terrain. Existing EOD vehicles can be easily adapted for the mission. We have modified a Hunter platform (figure 6), it is now equipped with a video camera, a metal mine detector fitted to the end of the robot arm and it has been interfaced with a PC.

The motion of the robot and the scan of the ground have been automated. When the metal detector detects something, the robot stops and an alarm is reported to the operator. Presently, we have used the original arm placed in front of the robot (Figure 6a) but the next step will be the development of a new scanner that will operate laterally (Figure 6b). This device and the robot will be controlled by a microcontroller of the same type as above; this will communicate via a RF serial link with the main control computer.



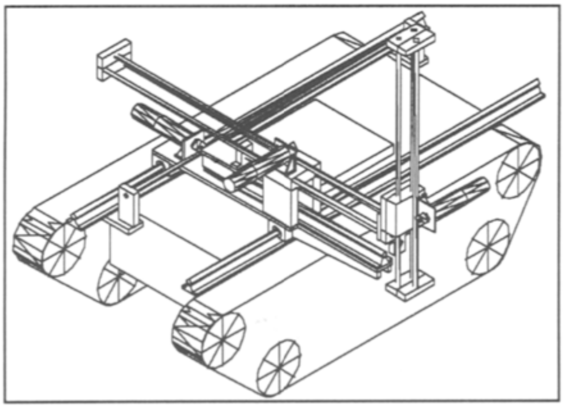


Figure 6. The original Hunter and the new scanning device

5. MAN MACHINE INTERFACE

In the case of a remotely controlled system and in order to efficiently control the system, the operator certainly needs the following elements :

- the video image of a camera placed on the robot,
- a symbolic representation of the mines field : a map must be automatically generated,
- the results of the data processing of the mines detectors.

The location of the robot in the field is essential if we want to control the robot remotely. First, if we use a lateral scanning we have to avoid the not yet examined ground and secondly we must be sure that the robot has explored all area. Several techniques are usable to locate a vehicle. We are presently working on the tracking of a color beacon put on top of the robot.

Knowing at each moment the position of the robot, we could generate a map with all detections. Besides, we also need to mark the position of the mines on the ground: this will be done with paint.

6. CONCLUSIONS

We don't pretend to solve the AP mines problem with our project, and perhaps our efforts are like a drop in the ocean, but we know that many people all over the world are working in the same direction. So, we are sure it will gives positive results in a not so far future.

7. REFERENCES

1. J. Weemaels, 1997, “Mines antipersonnel: état de l'art de diverses solutions de déminage”, Technical notes, Royal Military Academy, Belgium, 1997

2. Denis, Nauret, 1997, “Recherche de spécifications sur des plates-formes robotiques et des capteurs de détection de mines, dans le cadre d'un projet de déminage humanitaire”, Technical notes, Royal Military Academy, Belgium, 1997

3. Caprasse, 1997, “Les techniques de déminage”, Thesis report, Royal Military Academy, Belgium, 1997

4. *Proceedings of International workshop on Sustainable Humanitarian Demining*, SusDem'97, Zagreb, 1997

5. Y. Baudoin, “Development of low cost legged robots”,Proceedings of ISCMR'95, Slovakia; 1995

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Architecture for Embedded Systems in vehicles and mobile machinery

Bengt T. Henoch

Department of Electronic and Computer Systems
School of Engineering, Jönköping University
P.O. Box 1026, SE-551 11 Jönköping, Sweden

Abstract: Starting with a definition of embedded systems the paper reviews common architecture and some important background factors and important support systems influencing development trends and anticipated future architecture. Finally some R&D projects in process at Jönköping University are described.

1. Definition of embedded systems

It is well known that different products in an increasing way contain embedded systems, i.e. embedded eleectronics and associated soft ware.

Besides the wellknown IT and media products: computers, communication systems, radio, TV and others we find:

- ¤ White goods and domesticc products sucha as washing machines, refrigerators/freezers, stoves, lawn mowers, etc
- ¤ "Intelligent" houses and offices
- ¤ Leisure products such as toys, sports equipment, pleasure boats, etc
- ¤ Production and treatment equipment in manufacture, process and medicine
- ¤ Products from the transport and vehicle industry
- ¤ Mobile machinery in building industry, contract work, storage and freight terminals, agriculture, forrestry, etc

We will confine the discussion to embedded systems in vehicles and mobile machinery and there are several reasons for this:

- ¤ Vehicles and mobile machinery require well integrated embedded systems
- ¤ Vehicles and mobile machinery are parts of surface related application processes
- ¤ Vehicles and mobile machinery have embedded systems which distibuted over the whole product
- ¤ Vehicles and mobile machinery are trend setters

2. Typical architecture

Complex products are equiped with "embedded" electronics in the form of a controller and a "cluster" of sensors, actuators, displays and other types of inputs and outputs. Examples of such complex products are different types of machinery, consumer products and vehicles.

For the communication between controllers and "clusters" of inputs and outputs, different "very local" area networks have been developed and implemented. Driving in this development of area networks has been applications on board vehicles in automotive applications - vehicle area networks.

On board vehicles vehicle area networks are used for different applications and three classes of networks can be distinguished :

CLASS A Multiplex system for reduction of wiring between multiplex modules

CLASS B Multiplex system for eliminating os sharing redundant sensors and other elements

CLASS C Multiplex system for high data rate in real time control

The different networks have different applications and one vehicle can have a network architecture with several networks for:

- Dash board
- Communication units
- Climate control
- Electric lights
- Door and window management
- Motor management
- ABS

The different network islands can be connected to a common diagnostic bus, ISO 9141, for information exchange with an external diagnostic tester.

Presently, it can be said that four vehicle area networks have to some extent been specified and standardized. These networks, and derivatives from them, have found applications and support from different automotive industries.

The status of actual implementations of communication betwen controllers and "clusters" of inputs and outputs goes from staight forward wiring to the establishment of different control area networks e g for ABS and motor management and connecting gateways.

Below we give the basic parameters of the four vehicle area networks, which have common features in terms of an arbitration procedure on a multi-master network.

NETWORK	CLASS	SPEED kbit/sec	START bits	IDENT bits	DATA Bytes	ERROR detect	END bits
CAN(low)	B	10-125	1	12+6	0-8	CRC(16)	2+7
CAN(high)	C	<1000	1	12+6	0-8	CRC(16)	2+7
VAN	B	<125	1	12+4	0-28	CRC(16)	2
J 1850	B	10 / 41	1	24	0-12	CRC(8)	2
J 1939	C	250	1	12+6	0-8	CRC(16)	2+7
A BUS	C	500	1	12	2	REPEAT	2

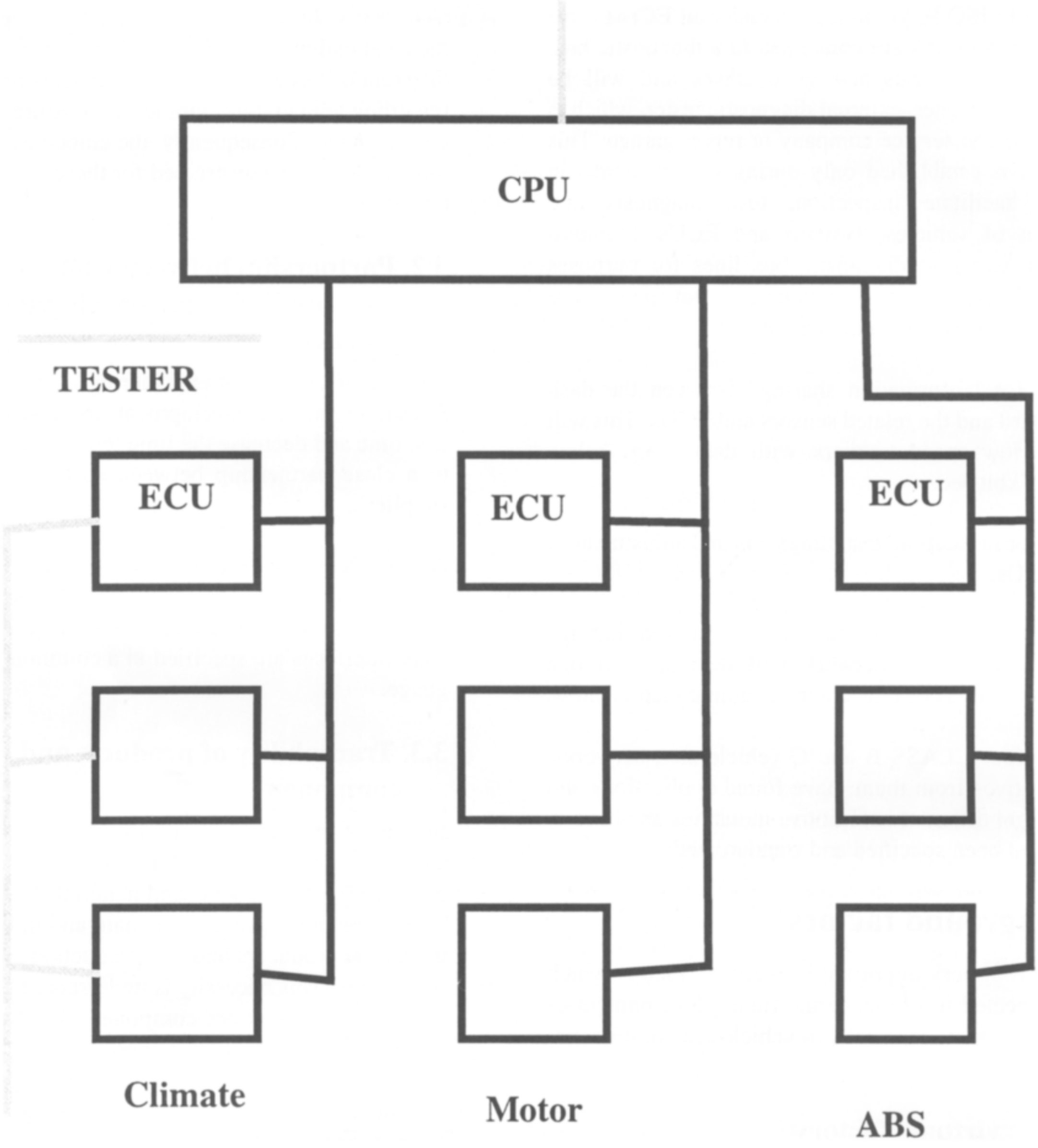


Figure 1. Typical architecture

Keyword Protocol 2000 is based on the standard ISO 9141 and ISO 9141-2 and described according to the OSI model in the three doccuments:

1 . ISO/CD 14230-1 Road vehicles -Diagnostic systems-Keyword Protocol 2000-Physical layer
2 . ISO/CD 14230-2 Road vehicles -Diagnostic systems-Keyword Protocol 2000-Data link layer
3 . ISO/CD 14230-3 Road vehicles -Diagnostic systems-Keyword Protocol 2000-Implementation layer

In automotive electronics different Electronic Control Units, ECU:s are having different functions like motor management, ABS management, climatic control, etc. The ECU:s are via a network connected to sensors and actuators and possibly to other ECU:s. The networks when used are taylored to their application and to existing time restraints. In the following we review some of the most common networks.

According to ISO 9141 standard usually all ECU-s have a diagnostic port and are connected to a diagnostic bus. The ECUs in the bus acts as a slaves and will be activated by a master, external diagnostic tester, which is connected at the service company or repair garage. This connection is established only during service works in order to facilitate inspection, test, diagnosis and adjustment of vehicles, systems and ECUs. Standard allows the using of diagnostic bus lines for purposes other than inspection, test and diagnosis, but special care shall be taken to avoid data collision.

- one for "information sharing" between the dash board and the related sensors and ECUs. This will be low speed network with data speed below 125kbit/sec.
- one for inspection, test, diagnosis and adjustment of ECUs.

The possibility also exists that the "information sharing" and the "diagnostic" network will develop into one general network as technical and economic factors allow.

Presently four CLASS B and C vehicle area networks, and derivatives from them, have found applications and support from different automotive industries and have to some extent been specified and standardized:

3. Background factors

There are four very important background factors which can be expected to influence the future development of embedded systems especially in vehicles and mobile machinery:

3.1. The "virtual" factory

With virtual factory is meant an organization where the principal producer and the subcontractors are connected by an IT network which supports all aspects from design, construction and production planning to quality control and product functionality. This gives requirements on the components and system modules that are delivered to the final assembly, that they must be complete including the embedded system and have well defined and at best standardized mechanical, electrical and soft ware interfaces for simple final assembly.

As a result the embedded systems move up to the front end of the process, and are subject to different difficult environmental conditions regarding temperature, vibration, moisture, corrosion etc. Consequently the embedded modules must be constructed for these conditions.

3.2. Partnership between customer and supplier in the development process

The use of concurrent engineering as a means for shortening the development and construction time and decrease the time to market leads to a close partnership between customer and supplier.

A side effect is that customer and supplier have common tools for construction, simulation and testing and that mechanical, electrical and soft ware interfaces are specified in a common language.

3.3. Traceability of products and components

In order to protect the environment and ensure adequate quality and maintenance it is necessary to components and material backwards in the production chain and link to individual producers and and production sequencies. This necessity is underscored by the requirement to trace components hurting the environment, through the complete production chain in time and space in all directions, by the demand for long term quality assurance through the monitoring of useage and performance data over the whole life cycle of components and products and by the advantage of a preventive maintenance, based on individual useage and performance data, during the whole life cycle. This type of traceability requires a significant storage of production and useage data, partly in central data bases and partly distributed in the embedded systems of components and products and an unambigous link betwen central and distributed data storage.

3.4. Mobile nodes in distributed IT systems

Increasingly vehicles and mobile machinery are equiped with support systems that link the mobile units to an external IT infrastructure. such systems are:

- * mobile communication for audio, data and video based on Mobytex, GSM, Inmarsat, etc
- * navigation and localization systems based on satellite ranging, inertial navigation, dead reckoning, etc
- * management and logistics systems for traffic, transport and surface related production processes systems for traffic, route, parking and service information
- * specific maintenance and monitoring systems for a population of didstributed objects, SCADA
- * security and alarm systems for hazarous goods, robery, accidents, rescue, etc
- * different combinations or functions for these systems

Presently these support systems generally are ad hoc and autonomous mutually and towards the mobile unit. The development is moving towards a tight integration of these systems and the embedded systems. Such an integration leads to an increased functionality e.g. regarding maintenance, transport logistics and finacial services and also decreased costs as redundancy between the systems is removed. A necessary condition for such development is that data security, authorization, concurrency and consistency have solutions with fire walls between the local systems and the different links to the infrastructure.

4. Important support systems

The influence of the background factors is demonstrated by the development of some very important support tools that at different hierarchical levels facilitate common interfaces between systems and HW/SW modules.

4. 1. Virtual components and devices

In the 80's a working group at General Motors developed a Manufacturing Application Protocol, MAP, which defined a common generic interface between a factory network and installed machine tools. The MAP protocol also specifies a Manufacturing Message Service, MMS, where a number of services are specified using C-routines.

The basic concept is that a typical machine or component is specified as the set of services that the machine can provide and the set of possible services are organized in a hierarchy of performance classes. A machine specified and classified in this way is called a Virtual Manufacturing Device, VMD, and is controlled through the network in the client/server concept.

For embedded systems the virtual device concept is applied to sensors, actuators, hydraulic components and other mechanical and electronic modules and a class of services are made available via a network interface. The network used is often CAN and the services are specified in a CAN Application Layer e.g. CAN Kingdom.

4. 2. Transparent operating systems for portable real time systems

A typical embedded system consists of a set of distributed HW/SW modules, ECU; which are communicating via one or several networks. The communication protocols are usually only specified up to layer 2 and for this reason higher communication protocols and network management functions must be implemented in SW in the HW/SW modules. This is a serious obstacle for the reuse and portability of application programmes.

For this reason the German and French automotive industry have specified a modular and open software architecture and corresponding network interfaces, OSEK/VDX. The architecture supports communication between units, network management and control of distributed realtime systems. OSEK/VDX compliance is starting to become a requirement for embedded modules in part of the automotive industry.

4,3. Gateway for external support systems

Offering state of the art entertainment, traveler information, communications, and security devices as parts of embedded systems in mobile units is hampered by the current design and by the difference in functional life time for heavy equipment versus IT systems. There is a need to install a variety of state of the art electronics devices after purchase, but installation can be complicated and expensive, and rarely results in integrated systems.

One problem is that manufacturers are using multiplex buses but . are not converging on a single standard. As a result, IT manufacturers must design and build multiple versions of their products to attach to the various buses, raising the cost. Another problem is that for warranty, safety, and liability reasons, electronic devices must not be allowed to interfere in any way with the operation of the mobile unit or of any of its subsystems, so they cannot be retrofitted to the multiplex bus.

A suggested solution is a dual-bus architecture connected to the auto's multiplex bus through a gateway. This allows electronic device manufacturers to build a single, automotive version of their product that plugs in to the external bus e.g. ITS Data Bus or USB In addition, such a Data Bus is independent of all vehicle systems, except for dc power. The gateway, under the control of the vehicle company, acts as a "firewall" allowing only authorized message traffic to pass between the vehicle multiplex bus and the Data Bus devices.

5. Development trends

Based on the reviewed background factors and developing support systems we will try and summarize the development trends. We will do this in a bottom up approach following the hierarchical levels of an anticipated future architecture:

1 . Components and modules such as motors, wheels, machine and gripping tools, transmission, steering and maneuvering, etc tend to be self contained units including embedded system and with complete mechanical, electrical and soft ware interface. As a consequence the embedded modules must be designed for a tougher environment.

2 . Components and modules will have a "virtual" interface with services organized in performance classes and defined in the application layer. Real time application progammes will be available based on virtual components.

3 . The internal network architecture will be more standardized with a continuation of a de facto standard based on different versions of CAN. Extended services for traceability, diagnostics and preventive maintenance will be available, cf KWP 2000

4 . The embedded system will tend to be distributed into subunits e.g. in road trains or in machine groups for surface related production processes and have shared resources especially concerning access to external services.

5 . Modular and open software architecture such as OSEK/VDX will become a requirement from major vehicle and machine manufacturers as a means to get portable and reuseable software and loss in real time will be compensated by increased speed and processing capability. The present trend of increasing embedded software size will be continued.

6 . The overall architecture will be a dual-bus structure with one bus for the internal system, one bus for connecting external support systems and a connecting gateway supporting also secure control and comand links

6. Future architecture

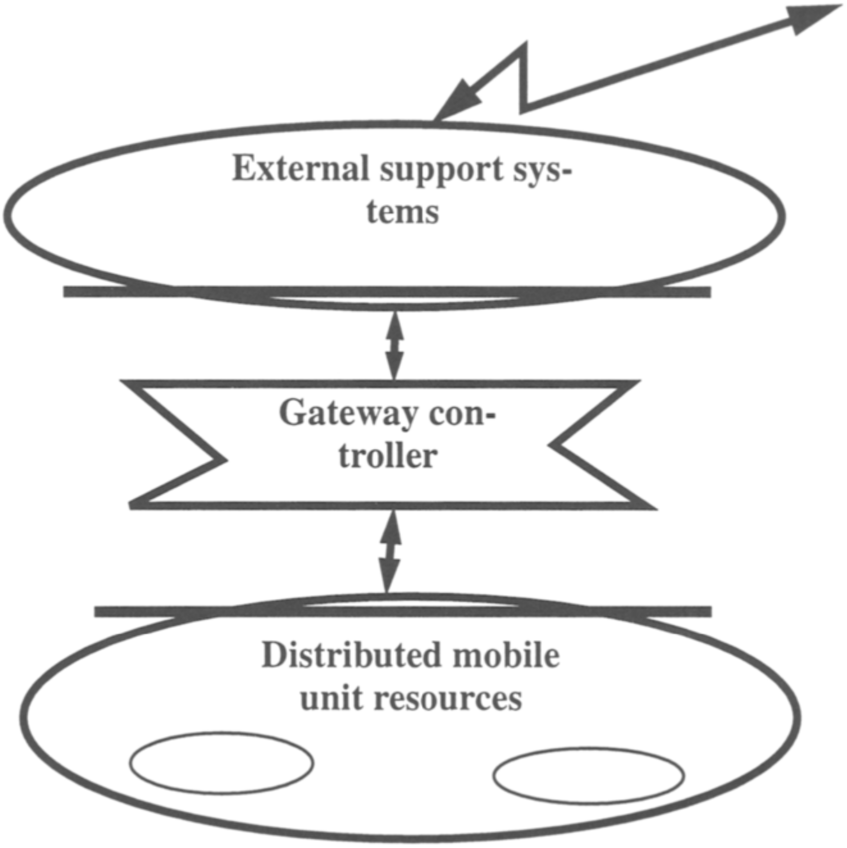


Figure 2. Future architecture

7. R&D at Jönköping University

At Jönköping University we are in the initial phase of an R&D programme that covers somme aspects of the anticipated future structure.

1 . Functional multi-chip modules for distributed motor management based on silicon carbide a material with potential to handle high temperatures, high currents and high voltages

2 . Linking of embedded system in differnt parts of road trains as a means for distributed fleet management and tracing.

3 . Requirements on linked information and external support systems e.g. GPS for integrated process chains in surface related production.

8. References

1 . General Motors MAP specification
2 .SSF 14230 Road Vehicles- Diagnostic Systems Keyword Protocol 2000 CAN OSI layers
3 . OSEK/VDX Open Systems in Automotive Networks, Workshop and Specification
4 . Strategic Needs and Future Trends of Embedded Software, TEKES Technology Review 48/96
5. SAE ITS Data Bus Committee, "In-Vehicle Electronics for IVHS - Workshop Results", November 30 & December 1, 1995
6. SAE ITS Data Bus Committee, "In-Vehicle Electronics for Intelligent Transportation Systems II - Workshop Results", July 9-10, 1996. 1

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Towards temporally embedded systems: dynamic short-term memory for mobile robots

R. M. Rylatt and C. A. Czarnecki

School of Computing Sciences, De Montfort University,
Leicester, LE1 9BH, U.K.

This paper identifies a problem of significance for approaches to adaptive autonomous agent research that seek to go beyond reactive behaviour without resorting to hybrid solutions combining symbolic and sub-symbolic AI systems. The feasibility of recurrent neural network solutions, including a novel architecture combining different existing recurrent neural networks, are discussed and compared in the light of experiments designed to test ability to handle long-term temporal dependencies.

1. INTRODUCTION

Much of the emphasis in behaviour-based research has been placed on embedding intelligent agents in space so that they would not suffer from the closed system brittleness of traditional, disembodied AI systems. Making run-time decisions based on deliberations actually prefigured at design-time has been a characteristic of such agents - for example [1]; typically they rely on some (usually inflexible) means to arbitrate between concurrently active modules, but temporal effects and dependencies were not modelled in any cognitive sense. Most connectionist research, too, has been concerned with spatial patterns or with giving essentially temporal structures, such as natural language, a spatial representation - a tendency that can be labelled *static connectionism*. Static systems are evidently closed in a similar sense to traditional AI systems in that they rely on a highly biased configuration of the input space before processing can take place effectively. Verschure [2] has argued that this approach amounts to giving the network a coded solution to the problem rather than the raw problem itself. On the other hand, situated, feedforward networks can deal with the temporal flow in the same way as behaviour-based systems. For example, limited obstacle avoidance can be achieved by training a feedforward architecture using a series of input vectors gathered by driving the robot or a simulation through a cluttered environment [3]. This is conventional real-time performance as displayed by spatially embedded systems. In a sense, however, truly open systems need to be embedded in time as well as in space, so that representations (however construed) are 'grounded' temporally as well as physically. Dynamical systems such as the recurrent neural networks (RNN) described below, provide answers to some of these practical difficulties and may promise a (very) long-term solution to the larger representational problems.

2. RELATED WORK

Examples can be found in the literature of connectionist systems that investigate the ability of physical or realistically simulated agents to perform temporally extended tasks such as sequential behaviour that require more than processing the current sensory pattern.

Yamauchi and Beer [4] evolved dynamic neural networks that have temporal processing power deriving from the ability of individual neurones to perform temporal summation. These nets were trained to reproduce sequences such as the regular expression {AB}* and more complicated sequences, but in this research the 'tasks' are only bits representing binary decisions idealised from an environment that is not simulated in detail. More commonly, networks that process through time are based on connections that feed back activations form earlier time steps. Ziemke [6] used second order recurrent networks (see section 3), and supervised learning to train a simulated Khepera robot to closely approximate finite state automaton behaviour while integrating obstacle avoidance - the behaviour demonstrated is essentially switching behaviour. Ziemke [5] compares second order recurrent networks (SORN) with the Elman network [9] and concludes that the former performs more robustly on these tasks. Tani and Fukumura [7] used very limited short-term memory STM for the recognition of place in order to make decisions at bifurcation points in a navigation task - a single step dependency was sufficient for their purposes and they at first used a tapped-delay line approach, though later, in [8] Tani adopted a Jordan-style network. Similarly, Pal and Kar [3] used a partial implementation of the Jordan model (omitting the self-recurrent context layer connections) to control a mobile robot in a navigation task with obstacle avoidance. In their paper there is some allusion to long-term dependencies but these are not analysed or investigated specifically and the task in fact proves to have only single step dependencies (the robot only needs to remember whether it was going forwards or backwards on the last time step).

2.1 The road sign problem: situated long-term dependencies

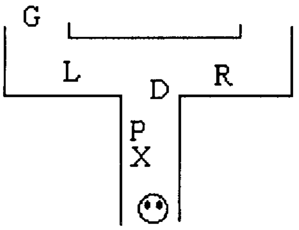


Figure 1: Sketch of the road sign problem showing the sign (X), concealed goal (G), decision point (D). The gap between X and D is represented by P (with acknowledgements to Claudia Ulbricht).

Ulbricht [10] describes how what she calls 'time-warped sequences' can be handled by a RNN - they are sequences in which significant elements are repeated an arbitrary number of times with arbitrary pauses between. For example, as sketched in figure 1, an agent might observe a particular environmental feature to its right on its way to a junction; it would continue to observe the feature in passing and then lose 'sight' of it, perceiving only the blank wall until it detects the junction ahead. When it arrives at the junction it must remember the significance of the previously observed feature so that it can decide which branch to take in order to reach a goal situated some further distance along one of the routes.

As we indicated at the outset, although it is couched in the terminology of autonomous agent research, the approach discussed above is closer to traditional connectionist experiments. These have a quasi-symbolic character (see [2]) and it is hard to see how such encoding schemes could scale even to a very simple simulated robot or animat with realistic sensors. Briefly, perceptual time slices are represented by symbols according to the current state, so that, for example, successive 'pause' states would be represented as {P*}. Obviously this is a very high-level abstraction of the real situation a situated agent would face. The novel recurrent network - the *input state network* - used in this work depends rather heavily on this kind of initial symbolic representation, subsequently using a method of transforming the input into a distributed representation. Obviously we would like our agents ultimately to form their own representations from raw sensory input as the basis for higher cognitive processes.

3. NEURAL NETWORK MODELS

The flexibility of the Elman network [9] makes it promising for our purposes. However, Ulbricht [10] appeared to have demonstrated that it did not cope with time-warped domains, though she conceded that an inappropriate training procedure might have been a factor in its failure rather than its inadequacy as a model for temporal processing. Initially, rather than discard an approach that has proved very effective in conventional connectionist research, it was decided to investigate the possibility of adapting it. Our reasoning was that the hidden layer could be induced to provide more input-relevant information for feedback as temporal context. The use of feedforward networks to compress input into hidden layer representations by training the network to reproduce its own inputs at the output layer (auto-association) is described in [11]. The Elman network was set up the to perform the auto-association task concurrently with the task of

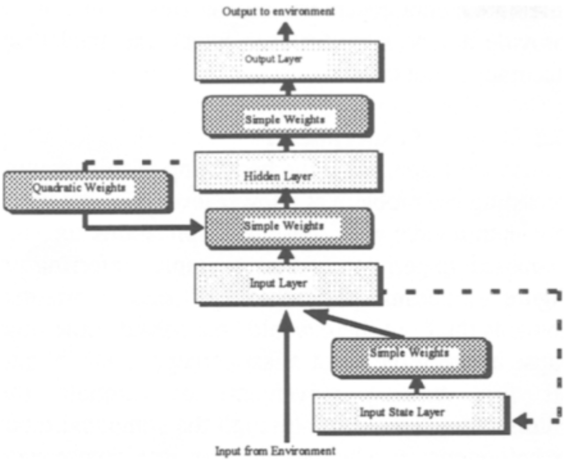


Figure 2: Hybrid architecture

associating percepts with motor outputs by adding extra output units equal to the number of inputs. In training, the inputs would be used as the target vector for the additional outputs so that a compressed input representation would form part of the hidden layer's state fed back as context. Additionally, in order to control the amount of state information that is recurrently processed, a constant gain was added to the context units. Thus the equation for the context layer's activation is:

$$c_j = (1-\gamma)h_j(t) + \gamma\, c_j(t-1) \qquad (1)$$

where, in equation 1, c_j is the j^{th} context unit, h is the j^{th} hidden unit and γ is the gain constant. The effect of the latter is governed by its reciprocal so that the sum of the state terms at time t and time t - 1 does not exceed unity, ensuring that the network does not saturate. If γ is set to zero, the effect is the same as that of the standard Elman context unit, that is, of a high definition, low-depth memory unit (using Mozer's terminology [12]). An increase in the gain will tune the context units by reducing the definition of the memory, a decrease will increase the depth of the memory. The idea is that the network will then be encouraged to generalise across temporal patterns in a manner analogous to the spatial generalisation observable in a feedforward network where a balance between the network's tendency to overfit functions, and its capacity to interpolate needs to be struck. At present this gain is set by hand, off line, but a self-adaptive gain is being investigated. Illustrated in figure 2 is the hybrid architecture we propose. It is based on the second-order or quadratic network described by Pollack [13] and the simple dynamic memory of Port et al, [14]. It also has similarities with the self-adapting recurrent network (SARN) of Ziemke [5] (a hybrid of a second-order network and the SRN). We leave an explanation of its task-oriented features until we describe the task itself - here we will outline its general aspects. Referring again to figure 2, it will be seen that this model has two levels of feedback (dotted arrows). Firstly, feedback occurs from the input layer to itself - we have indicated this by showing an extra 'layer' at a level before the input layer, similar to the context layer of an SRN. We believe that this re-conceptualisation may blur the distinction between 'layer feedback' and 'unit feedback' drawn in [10], but in fact it shows the fully interconnected (solid arrows) nature of the self-recurrent input layer quite clearly, conforming to the activation equation in [14]:

$$y_i(t+1) = squash[\alpha y_i(t) + \sum w_{ij} y_j + input + bias \quad (1)$$

where α is a decay rate, y is the feedback and *input* represents direct sensory input. The second level of feedback is from the hidden layer and this incorporates the second-order aspects. We show the architecture as a single network but it functions as two subnets in a master-slave relationship. Feedback, from the hidden layer at time $t-1$, is construed as the input layer of the master network, which is then fed forward through a linear function to its output layer. The units of this output layer, in our model, represent the weights of the slave net hidden layer. The terms 'quadratic' and 'simple' reflect the multiplicative order of the connections. It will be seen that the master subnet will have a number of weights equal to the multiple of the simple weights between the layers it connects and the number of units in its input layer. The model is trained using BP in the manner found to be successful by Elman [9] and Pollack [13] which effectively backpropagates error one step in time. In this case, as in [6], the hidden layer weight vector, adjusted by the slave net at time t - 1, serves as the target vector for the master net at time *t*. Thus, the master weights multiplex the activation function of the slave hidden layer, giving (from our perspective) richer representational potential.

3. EXPERIMENTS

3.1 Fine grained simulation

For these experiments, we used an integrated mobile robot and neural network simulator (IMRANNS). This was designed and implemented primarily for the development and testing, in fine-grained simulation, of modular neural network mobile robot control architectures - see, for example [15].

Although our aim was to cast the problem at a much lower level of abstraction, our simulated environment is still relatively impoverished because the robot has limited perceptual abilities and the environment is two-dimensional. Nevertheless, we believe that similar simulations have transmigrated successfully to real robot demonstrations, according Pal and Kar [3]. For our experiments, we have used an idealisation of active sonars as range finders, similar to the description in [16]. The latter provided adequately realistic, approximate distance information for landmark recognition from a cognitive sketch map, although we employ an original scan line algorithm rather than Bresenham's algorithm. A similar sonar model can be found in [3], mentioned above. More accurate range profiles, of the general kind we model, would doubtless be obtainable, if necessary, using laser range finders of the kind used by [7]. So there is no reason to suppose our simulation is unrealistic in its sensory demands. Additionally we use the robot's simulated photoreceptors, but as this is merely to provide a 'go' stimulus (see later), the modelling accuracy is not critical.

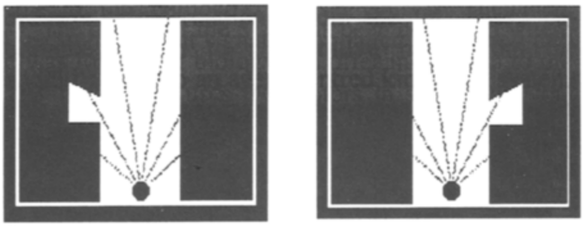


Figure 3: Mirror image environments (first experiment)

3.2 Design of experiments

As we suggested at the beginning of this section, recasting the problem at a lower level of abstraction has entailments with design implications for the proposed experiments. For example, referring to figure 2, because the robot has sensors oriented towards the front there would be a considerable time lapse between the robot's last perception of 'D' and its first turning movement to negotiate the intersection. This is so although the simulated robot is holonomic; non-holonomic types would pose even greater problems. The long dependency (in neural network terms) of these temporal events presents a difficult problem - for example, even the non-situated input state network (see above) seems to be reliable only up to around six time steps into the past. Figure 4 shows the mirror image environments used for training and testing in the first set of experiments. In this the robot perceives the instructional stimulus (a 'notch' in the 'solid' wall to delimit a zone that should give a distinctive sonar or lasar range finder profile in contrast to the flat expanse on either side). In the first experiments, the 'sign' was located either to the right or to the left of the central aisle. The delayed response task is to execute a 'U' turn, the direction being determined by which side the stimulus was perceived. There is no 'go' signal - the robot simply learns and executes the complete tour, so in these experiments only long-term dependencies can be investigated. Apart from the reflection about the vertical axis of symmetry, these two environments are identical in all respects. This is to ensure that the neural network does not distinguish accidentally unbalanced features, and use these as the basis for its decisions, rather than its memory of the distinct range-finder profiles. We took the same precautions for the second set of experiments, designed to investigate both long-term dependencies and generalisation across time combined with a more demanding recognition object as the stimulus. In these, the instruction stimulus is perceived as one of two distinct range-finder profiles (each a distinct shape on the same side of the aisle) in another pair of environments. A 'go' signal was given to the

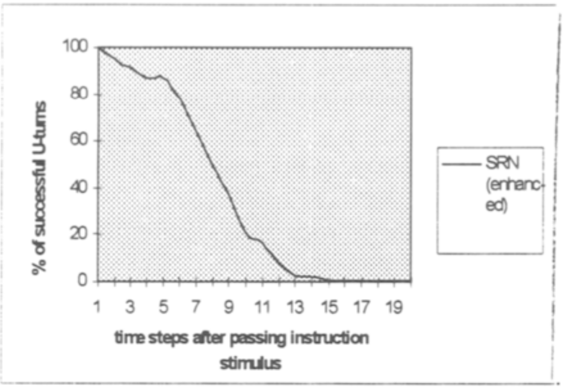


Figure 4: First set of experiments

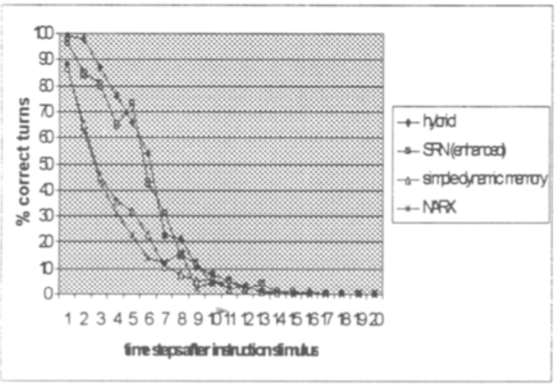


Figure 5: Second set of experiments

robot by means of a light beam. We added more range finders to the robot for this task.

The training procedure used was as follows: continuous samples of the robot's sensory inputs and corresponding motor responses were collected automatically as the robot was moved manually along the centre of the aisle in each of the two environments. If the experiment did not involve a 'go' signal, after a fixed number of steps following the instruction stimulus has disappeared from 'view', the robot was turned about its axis. In this case, the total number of steps before the turning point was varied across multiple training sets. When there was a 'go' signal this was incorporated in the training set - the light was 'flashed' for one time step at a given number of steps after extinction of the stimulus. The network was then trained on each training set, and tested on the robot in each environment respectively to confirm that the appropriate behaviour had been learned in response to each distinct stimulus. As expected, learning this behaviour proved to be a relatively trivial problem for all the nets tested.

The procedure for training on the overall task was initiated by concatenating the two training sets obtained from the paired environments for a given trial. The network was then trained using this expanded training set.

After training the robot was tested in each of paired environments and its ability to perform the required delayed response was assessed.

3.3 Results

Results, averaged across trials for the first experiment, based on just the enhanced SRN (described above) as subject, are summarised in the chart in figure 5. As there is a 50 per cent chance of turning in either direction, the results have been re scaled to set chance-level performance at zero. Clearly this SRN has some ability to handle long-term dependencies in a continuous domain - cf. the findings in [10] showing the complete inability of the SRN network to accommodate these in a near-traditional connectionist 'environment'.

The second set of experiments involved several different recurrent models. These comparative results are summarised in figure 5, again re-scaled as above. We found that these results largely confirmed our expectations that input level information is most useful to the agent, but the reasonable performance of the NARX [17] network (with three output feedback delay lines) is somewhat surprising and merits further investigation.

5. CONCLUSIONS AND FURTHER WORK

We have begun to articulate the kind of problems that in our view must be solved before agents with genuine, autonomous deliberative powers can emerge. It is our belief that some form of STM that is more than an inflexible spatialisation of temporal events is a *sine qua non.* Our experiments show how simple algorithms can be used in conjunction with recurrent neural network architectures to achieve observable behaviours that have some characteristics looked for in paradigmatic STM experiments. Our argument concerning the significance of these ideas for foundational AI concerns has been more fully developed in [18]. We hope to report more detailed, statistical analysis of the results of these and further experiments in another extended paper.

REFERENCES

[1] Brooks, R.A. Intelligence without representation, *Artificial Intelligence*, 47, 139-159, 1991.

[2] Verschure, P.M. Connectionist explanation: taking positions in the mind-brain dilemma. In: Dorffner, G. (Ed.) *Neural Networks and a New Artificial Intelligence*, International Thompson Computer Press, 134-187, 1993.

[3] Pal, P.K. and Kar, A. Sonar-based mobile robot navigation through supervised learning in a neural net. *Autonomous Robots*, 3, 355-374, 1996.

[4] Yamauchi, B. and Beer, R., Integrating reactive, sequential and learning behaviour using dynamical neural networks. *From Animals to Animats 3: Proceedings of the Third Int. Conf. on the Simulation of Adaptive Behaviour,* 383-394, MIT Press, Cambridge, M.A., 1994.

[5] Ziemke, T. Towards adaptive behaviour system integration using connectionist infinite state automata. *From Animals to Animats 4: Proceedings of the Fourth Int. Conf. on the Simulation of Adaptive Behaviour*, MIT Press / Bradford Books, Cambridge, M.A., 1996.

[6] Ziemke,T. Towards adaptive perception in autonomous robots using second-order recurrent networks, *Proc. of the First Euromicro Workshop on Advanced Mobile Robots*, Kaiseslautern, 89-98.

[7] Tani, J. and Fukumura, N., Learning goal directed sensory-based navigation of a mobile robot. *Neural Networks.* Vol. 7, No, 3. 55-564, 1994.

[8] Tani, J. Model-based learning for a mobile robot navigation from the dynamical systems perspective. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, Vol.26, No.3, June 1996.

[9] Elman, J. Finding structure in time, *Cognitive Science*, 14, 179-211, 1990.

[10] Ulbricht, C., Handling time-warped sequences with neural networks. *From Animals to Animats 4: Proc. of the Fourth Int. Conf. on Simulation of Adaptive Behaviour*, 180-192, 1996.

[11] Kadaba, N., Nygard, K.E., Juell, P.L. and Kanga, L., Modular backpropagation network for large domain pattern classification. *Proc. of IJCNN*, Washington D.C., Vol. II., 551 - 554, 1990.

[12] Mozer, M.C. (1993). Neural net architectures for temporal sequence processing, In: A.Weigend and N.Gerschenfeld (eds.), *Predicting the Future and Understanding the Past*, Addison Wesley.

[13] Pollack, J.B. The induction of dynamical recognisers. In: Port, R.F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition.* MIT Press, Cambridge, MA, 1995.

[14] Port, R.F., Cummins, F. and McAuley, J.D. Naive time, temporal patterns and human audition. In: Port, R.F. and van Gelder, T., editors, *Mind as Motion: Explorations in the Dynamics of Cognition.* MIT Press, Cambridge, MA., 1993.

[15] Rylatt, R. M., Czarnecki, C. A. and Routen, T. W., A neural network based adaptive controller for mobile robots. *Proc. of the Twelfth Int. Conf. on Systems Engineering*, Coventry, U.K., Vol. 2, 587-592., 1997.

[16] Stein, L.A. Imagination and situated cognition, In: K.M. Ford, C.Glymour and P.J. Hayes (eds.) *Android Epistemology*, AAAI Press/Press, Menlo Park, CA,1995.

[17] Lin, T., Horne, B.G., Tino, P. and Giles, L., Learning long-term dependencies is not as difficult with NARX recurrent networks, *IEEE Transactions on Neural Networks*, 1996

[18] Rylatt, Mark, and Czarnecki, Chris. Beyond physical grounding and naïve time: investigations into short-term memory for autonomous agents. In: *From Animals to Animats 5, Proceedings of the Fifth Int. Conf. of the Society for Adaptive Behaviour*, Bradford Books, MIT Press (in press).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Automatic guided vehicle with independent rear wheel dc motor drivers and front wheel steering

V. González [a], J. Guadarrama [b] and M. López [c]

[a,b] Departamento de Ingeniería Mecatrónica, Facultad de Ingeniería, Universidad Nacional Autónoma de México,
Circuito Exterior, Ciudad Universitaria, México D.F., 04510, México
vjgv@servidor.unam.mx[a], jgl@servidor.unam.mx[b]

[c] Centro de Diseño y Manufactura, Facultad de Ingeniería, Universidad Nacional Autónoma de México,
Circuito exterior, Ciudad Universitaria, México D.F., 04510, México.
lopezp@servidor.unam.mx

Abstract: This paper presents the different issues concerning the design of an Automatic Guided Vehicle (AGV) which features two driving motors for the rear wheels and one steering motor for the front wheel. This configuration allows to have a short turning radius. The AGV incorporates, in its first stage of the development, an infrared light as guided system. Keywords: AGV, Mechatronics, mobile robotics, kinematics, dynamics, control.

1. NOTATION

1.1 Kinematics modelling

b : Distance between the rear wheels
C : Middle point of b
θ_s : Turning angle of front wheel
r_1 : Turning radius of rear wheel 1
r_C : Turning radius of point C
r_2 : Turning radius of rear wheel 2
R_s : Turning radius of front wheel
L : Distance between rear and front axles
V_1 : Tangential speed of rear wheel 1
V_2 : Tangential speed of rear wheel 2
V_C : Cruising speed of point C
w_1 : Angular speed of rear wheel 1
w_2 : Angular speed of rear wheel 2
e : Sensor arm length
d : Angle between front wheel and sensor arm

1.2 Dynamic modelling

F_t : Traction force
F_{ft} : Force opposing rear wheel rolling
F_{fd} : Force opposing front wheel rolling
F_f : Force opposing equal wheels rolling
F_i : Inertial force
F_a : Air friction force
W_m : Vehicle weight
m : Total mass of the vehicle
T_t : Total wheel torque
T_m : Motor torque
a : Linear acceleration of the vehicle
α_L : Angular acceleration of the wheels
α_m : Angular acceleration of the motor
g : Gravity acceleration
w : Angular velocity of the wheels
J_e : Total inertia at the motor shaft
I_{rd} : Front wheel inertia
I_{rt} : Rear wheel inertia
I_r : Inertia of equal wheels
I_n : Transmission inertia (n-th element)
r_d : Wheel radius
i_n : Transmission speed ratio (n-th element)
i_t : Total transmission speed ratio of the system (w/w_f)

η : Total efficiency of the transmission (0.98 gear transmission)

1.3 Field constant dc motor modelling

S : Laplace operator
R_a : Electric resistance
L_a : Inductance
K_m : Torque constant
Kb_m : Back electromotive force constant of the motor
B : Viscosity friction coefficient
J_e : Total inertia at the motor shaft
I_a : Electric current
e_a : Applied voltage at the motor terminals
e_b : Voltage of the back electromotive force
w_f : Angular speed of the motor shaft
T_m : Motor torque

1.4 Selection and tuning of the control mode

H (s) : Motor transfer function
G (s) : Feedback transfer function (tachometer)
C (s) : Transfer function of the mode control
Ho(s) : Total transfer function of the control system
V_{ref} (s) : Reference voltage of the system
K_p : Proportional gain
T_i : Integration time constant
K : Motor gain
T : Time constant of the motor
K_{fin} : Total gain of the control system
T_{fin} : Total time constant of the control system
Kb_g : Back electromotive force constant of the tachometer

2. INTRODUCTION

Automatic Guided Vehicles (AGV's) are commonly used as material handling and transport devices with free moving and flexible routes within Flexible Manufacturing Systems (FMS). In general, rail guided vehicles have better performance than autonomous vehicles incorporating obstacle evasion algorithms. The main advantage of vehicles that navigate tracking a floor reflective tape is their inherent flexibility and their low cost of installation on the plant.

This paper presents the design of an AGV that uses two driving dc motors for the rear wheels and one wheel in the front to steer the vehicle. The design allows to have a short turning radius that depends only on the mechanical turning limitations of the front wheel and the distance between the front and rear axle. If the middle point along the rear axle is taken, then the turning radius could be zero. The current project is being developed at the Mechatronics Department, UNAM, and it is an example of mechatronic design.

2.1. Power propulsion system

The power propulsion system is based on two independent dc motor drivers, as shown in fig 1. The angular speed algorithm of both wheels is based on a kinematic model of the AGV. The speed feedback control system allows to obtain the precise angular speed needed for each wheel. The kinematic restrictions are met and wheel slip is prevented.

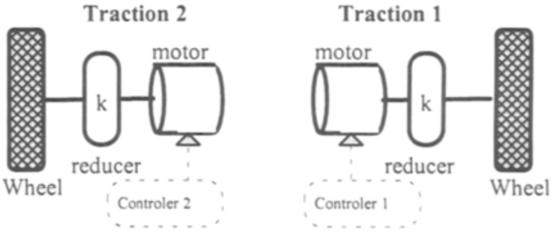


Figure 1

2.2. Steering system

The steering system is based on one front-wheel steering mechanism, as shown in fig. 2.

2.3. AGV guiding system

The AGV guiding system is based on the transmission and reception of an infrared light. The infrared light is reflected on a floor reflective tape. A single or double sensor arrangement can be used in the vehicle. The latter is more reliable. The one sensor tracking control algorithm continuously oscillates around the track. The one sensor system works as follows : the sensor detects the reflection of the track, the steering turns to find the unreflected area, then when the sensor stops receiving the signal, the steering system turns back to find the reflective tape again. The double sensor arrangement does not present this oscillatory behaviour when the AGV is navigating. Some oscillation is produced during tracking transitions, and when the route direction or tuning radius changes. The strategic positioning of the sensors is crucial for achieving a minimum turning radius since the steering wheel must always be tangent to the route, as seen below.

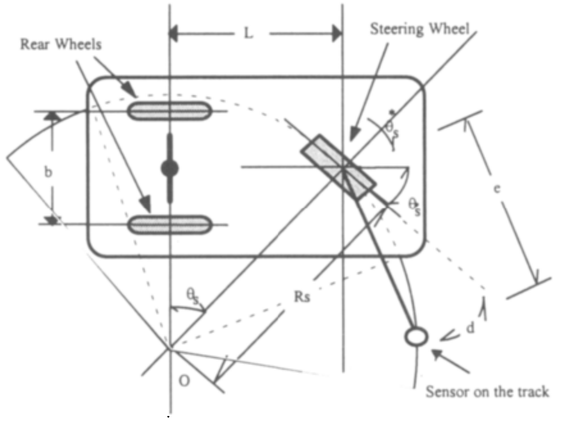


Figure 2

Using trigonometric relations, the d angle can be obtained, as shown in equation (1) and fig. 3, when y=e / L and x = θ_S. Since the total sensor angle is θ_S plus d then the angle of the wheel must be adjusted for each total sensor angle to keep the AGV on the track.

$$d = 90 - 2\ \mathrm{ArcTan}\left[\frac{2 - y\ \mathrm{Sin}[x]}{2 + y\ \mathrm{Sin}[x]}\right] \qquad (1)$$

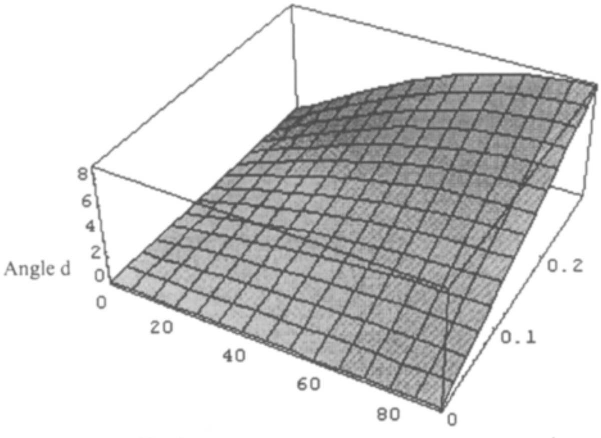


Figure 3

2.4. Communication system

The communication between the AGV and both the central station and work stations is base on a low range bidirectional infrared light system. When the AGV needs to get information from the system it must stop by any station and then wait to receive the required instructions.

2.5. Battery system

The AGV power system has a bank of rechargeable batteries. The batteries charge is carried out while the AGV is out of work in the recharge station.

2.6. Control system

The AGV control is based on a Motorola MC68HC11 micro-controller who takes control of all the actions of the system and the logistics of the process. It also interacts with the distributed PI angular velocity controller of each rear wheel.

3. KINEMATIC MODELLING

Diverse vehicle configurations can be found in the literature [Zaho, 1992]. The automatic guided vehicle with three wheels, two-rear-drive-wheel and front-steering-wheel is shown in fig. 4. The minimum turning radius (R_S) depends on the turning limitations of the front wheel (θ_S) and the distance between the front and rear axle (L). It is important to emphasise that if the rear wheels spin in opposite directions and the θ_S angle is ± 90^0, then the turning radius r_C can be zero and the AGV rotates about point C. From the analysis of fig. 4, we obtain the angular speed of each rear wheel as a function of b, L, V_C and θ_S.

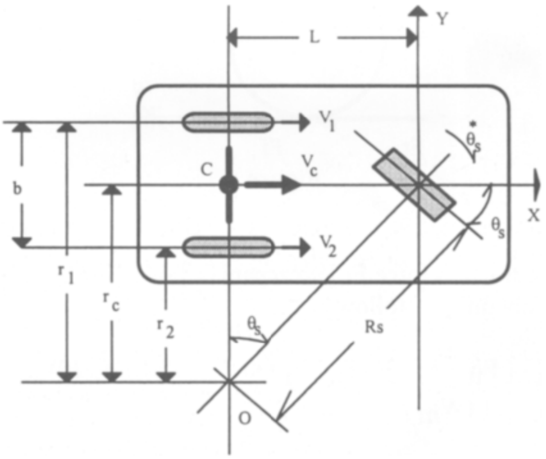


Figure 4

$$r_c = L / \tan\theta_s \tag{2}$$
$$R_s = L[\ 1 + (\ 1/\tan\theta_s)^2\ ]^{1/2} \tag{3}$$
$$V_1 = w_1 r_1 \tag{4}$$
$$V_2 = w_2 r_2 \tag{5}$$
$$V_c = (V_1 + V_2) / 2 \tag{6}$$

Each wheel velocity, are :

$$V_1 = V_c\,[1 + (b/2L) \tan\theta_s\,) \tag{7}$$
$$V_2 = V_c\,[1 - (b/2L) \tan\theta_s\,) \tag{8}$$

4. DYNAMIC MODELLING

The control tuning selection depends on the total inertia reflected on the motor shaft (J_e). The dynamic analysis (fig 5 and 6) to obtain the total inertia reflected on the motor shaft is presented below.

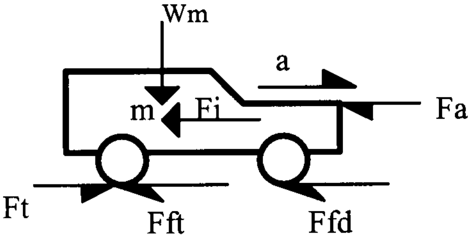


Figure 5

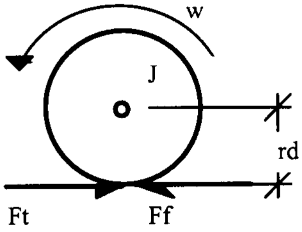


Figure 6

the traction force Ft [Szczepaniak] of figure 5 and 6 is obtained as follows:

$$F_t - [\,F_{ft} + F_{fd}\,]\ - F_a - F_i = 0 \tag{9}$$
$$Fi = a\,[\,(W_m/g) + [\,I_{rd} + I_{rt} + \Sigma^n_{i=1}(I_n/i_n^2)\,]\,(1/r_d^2)\,] \tag{10}$$

with the following considerations :

$$\text{Velocity} < 40 \text{ km/hr} \quad \Rightarrow\ F_a = 0 \tag{11}$$

Equal wheels =>

$$F_{fd} = F_{ft} = F_f \tag{12}$$
$$F_f = W_m\ f \qquad (0.010 < f < 0.017) \tag{13}$$
$$I_{rd} = I_{rt} = I_r \tag{14}$$

$$W_m = mg \tag{15}$$
$$F_t = Tt / r_d \tag{16}$$
$$T_t = T_m\,P(\,\eta / I_t\,) \tag{17}$$
$$T_m = J_e\,\alpha_m \tag{18}$$
$$a = \alpha_L\,r_d \tag{19}$$
$$\alpha_L = \alpha_m\,i_t \tag{20}$$

from (9) to (20) and ignoring the transmission inertia $\Sigma^n_{i=1}(\,I_n/i_n^2\,) = 0$ we conclude that:

$$J_e = [i_t^2 / \eta\,]\,[\,m\,r_d^2 + 2\,I_r\,] \tag{21}$$

5. FIELD CONSTANT DC MOTOR MODELLING

The field constant dc motor modelling, is presented below:

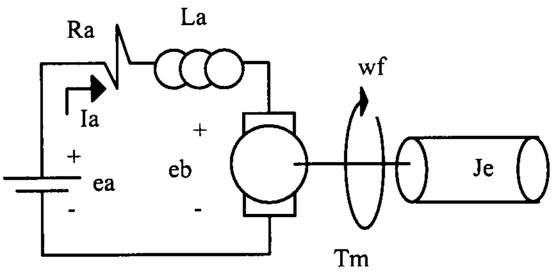


Figure 7

$$I_a = [e_a - e_b] / [L_a S + R_a] \tag{22}$$
$$T_m = K_m\,I_a \tag{23}$$
$$w_f = T_m\ /\ (SJ_e + B) \tag{24}$$

$$e_b = Kb_m w_f \qquad (25)$$

with the following considerations

$$L_a \Rightarrow 0$$
$$B \Rightarrow 0$$

the transfer function is :

$$H(S) = \frac{W(S)}{e_a(s)} = \frac{K}{TS + 1} \qquad (26)$$

when K is the total gain and T is the time constant of the motor :

$$K = \frac{1}{Kbm} \quad ; \qquad T = \frac{Jt\,Ra}{Km\,Kbm}$$

6. SELECTION AND TUNING OF THE CONTROL MODE

Using the basic scheme or feedback controls shown in fig. 8 we select the control mode and perform its tuning. Many tuning methods exist [Astrom, 1995].

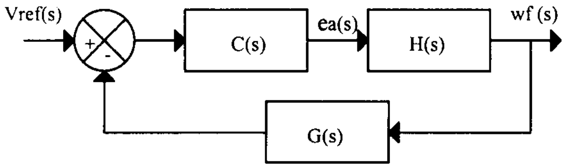


Figure 8

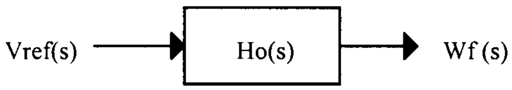


Figure 9

Since we know H(s) y G(s), we can get the transfer control function C(s) who satisfies the desired transfer function Ho(s) of the total system (see figure 9). The transfer function Ho(s) can be selected according to the needed system response, that is, from fig. 8 :

$$Ho(S) = \frac{Wf(s)}{Vref(s)} = \frac{C(s)\,H(s)}{1 + C(s)H(s)G(s)} \qquad (27)$$

in function of C(s) :

$$C(s) = \frac{1}{H(s)}\,\frac{Ho(s)}{1 - G(s)\,Ho(s)} \qquad (28)$$

if Ho(S) satisfies the following transfer function :

$$Ho(s) = \frac{Wf(s)}{Vref(s)} = Kfin\,\frac{1}{Tfin\,S + 1} \qquad (29)$$

when $\mathbf{K_{fin}}$ is the total gain and $\mathbf{T_{fin}}$ is the total time response wanted.

Substituting (26) , (29) and $G(s)= Kb_g$ we get C(s) as a PI control:

$K_p = (K_{fin}/T_{fin})\,(T/K)$ (30)

$T_i = T$ (31)

$K_{fin} = 1/\,Kb_g$ (32)

that is :

$$C(s) = Kp\left[1 + \frac{1}{Ti\,S}\right] \qquad (33)$$

7. AGV OPERATION

Once the dimensions of the AGV have been defined and the vehicle is on the track, the cruising speed (V_C) and the θ_S+g angle become the input variables. The main internal algorithms are two. The first consisting in getting the θ_S angle when the vehicle follows the track, as mentioned in point 2.3. The second one computes the angular speed $w_1=V_1/r_1$ and $w_2=V_2/r_2$ that are functions of the cruising speed V_C and the θ_S angle, as described in section 3. The voltages $Vref_1$ and $Vref_2$ can be got from w_1 and w_2 using $V_{ref} = (1/K_{fin})\ w_f$ from (29) and are provided by the DAC's of the microprocessor system. The control mode of each wheel can be analog or digital, and the tuning of the system will depend on the K_{fin} and T_{fin} of the total control system (as discussed in 6).

8. CONCLUSIONS

Mechatronic design includes the use of mechanics, electronics, and software. In this work it was evident that the application of the systems modelling theory, control and signal processing is also required. The AGV features a minimum L turning radius for R_S. In *the middle point between rear wheels the turning radius can be zero* if the drive motors can be powered in reverse. This is useful when the AGV is used as a free mobile robot using other guided system as a Vector 2X Compass Module and a Positioning System by Triangulation.

REFERENCES

1. Astrom, K, and Hagglund, T, PID CONTROLLERS: THEORY, DESIGN AND TUNING, 2nd Edition, Instrument Society of America (ISA), USA, 1995.
2. Bernal, F., Sousa, R, Gonzalez, V, DISEÑO Y CONSTRUCCIÓN DE UN VEHICULO DE GUIA AUTOMÁTICA PARA UN SISTEMA DE MANUFACTURA FLEXIBLE, Tesis de Licenciatura, Facultad de Ingeniería , UNAM, CU, México D.F., 1993.
3. Burr, J. MECHATRONICS DESIGN IN JAPAN, Institute of Engineering Design, Technical University of Denmark, Denmark, 1989.
4. Burr, J. A THEORETICAL APPROACH TO MECHATRONICS DESIGN, Institute of Engineering Design, Technical University of Denmark, Denmark, 1990.
5. Ogata, K, INGENIERIA DE CONTROL MODERNA, Segunda Edición, Prentice-Hall Hispanoamericana, México, 1993.
6. Szczepaniak, C., Aragón, R. FUNDAMENTOS DE DISEÑO DEL AUTOMOVIL, Ed.CECSA, México, 1982.
7. Gonzalez,V, Guadarrama, L, CONSIDERACIONES BASICAS EN EL DISEÑO DE UN VEHICULO DE GUIA AUTOMATICA CON DIFERENCIAL ELECTRICO, Memorias del III Congreso Anual de la Sociedad Mexicana de Ingeniería Mecánica A.C., México D.F, México, 1997.
8. Zhao, Y. and BeMent, S.L., KINEMATICS, DYNAMICS AND CONTROL OF WHEELED MOBILE ROBOTS" Proceedings of the 1992 IEEE, International Conference on Robotics and Automation, pp. 91-96, Nice, France, May 1992.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

An Intuitive Control and Sensory System for the Enhancement of Remotely Controlled Vehicle Operation.

Dr. P. Webb and Dr. J.I. Robson.

Department of Manufacturing Engineering and Operations Management
The University of Nottingham, University Park, Nottingham, NG7 2RD.

This paper describes some of the key elements of a Remotely Controlled Vehicle operating system that has been designed using human factors design principles. The described system is for a land vehicle but the generic methods and systems described would be equally applicable to airborne and under sea vehicles.

1.Introduction

Remotely controlled vehicles (RCVs) are finding applications in the air, on land and beneath the sea. On land the primary areas of deployment are in hazardous and unpleasant environments performing tasks such a explosive ordinance disposal (EOD), mine clearance, nuclear decommissioning[1] and for remote operations in the petrochemical industry. In the air, remote vehicles are used for reconnaissance and as target drones. Under the sea, vehicles are used for surveying, exploration and in the offshore oil industry. Whatever the operating medium most of these systems are controlled by a human operator located at some remote control centre. Control signals are fed to the vehicle along a link which may be cable, fibre optic or radio, whilst sensory information and/or television pictures are returned to the operator. The final effectiveness of such systems is limited not only by the technological complexity of the control system, but also by the ease with which the operator can interact with the control system and how intuitive the required operator inputs are. Of particular importance in assessing the effectiveness of such a system is the degree of closeness with which the users mental perception of the configuration and status of the vehicle matches that of the real system.

2. Human Factors Design Principles

There is always a danger with increasing technological sophistication that operators will become overburdened by a myriad of control systems, and information sources ; potentially resulting in a deterioration in task performance. The operator of a remotely controlled vehicle may be responsible for the control of a range of different devices and systems, including light, cameras, grippers, weapons etc. Furthermore, the guidance and navigation of the vehicle itself may place considerable demand on the operator's cognitive resources.

The tasks that robotic systems perform can be very complex, requiring a great deal of skill, training and experience. The levels of autonomy provided to the human operator are likely to be determined by economic constraints, and technical limitations. It therefore seems likely that the role of the human operator is safe for some time to come[2]This is especially pertinent to operations involving a remote control vehicle in unstructured domains. Remote worlds can be highly structured and static. Rules of location, geometry, and orientation mean that it is relatively easy to model this world. Unstructured worlds lie at the opposite end of this continuum. In the unstructured world, the operator may have no advance knowledge of the terrain, location and size of targets etc.

2.1.Telepresence in the unstructured world.

Where pre-programmed operations are infeasible, the operator is required to act in a supervisory role because of operational Control systems and interface features must ensure that conceptual models of tasks are utilised in the design of complex interactive systems.

The development of an appropriate conceptual model is essential for the efficient use of any tool or piece of equipment, regardless of whether it is a domestic CD player, vehicular cruise control, or the teleoperation of a semi-autonomous robot. The key to successful conceptual modelling is understanding the image that is conveyed to the operator. The image that the operator develops is drawn from the physical appearance of the overall system, (the system image) from the RCV structure, the control systems which are used to manipulate it, and the associated input/output devices. These all contribute to the conceptual model. It is essential therefore that care is taken during design, in order to ensure that the an appropriate conceptual image about how the system works is conveyed to the operator. If there is a lack of compatibility and consistency between the system image and the operator's conceptual model then there will be problems associated with its use.

It has been argued that psychological variables differ from physical variables.[3] Failure to understand and overcome the difficulties represented by these differences will result in poorly designed interfaces. This is integral to a cognitive engineering approach to systems design. So what does this mean in terms of telepresence?

Remote operations in unstructured environments may be hazardous and stressful for the operator. It is essential therefore that systems are engineered in such a way that operator performance is safe and efficient.

The RCV system that we are working on requires control and navigation in remote, unfamiliar, and potentially unstructured environments. The attributes which make up the conceptual model for the user are more complex because of the interface elements which make up the human-interactive computer system[4]. (HIS). (Level 1 in Figure 1), and the Task-Interactive System (Level 3), which manipulates the sensors and actuators (Level 4), to carry out the task remotely.

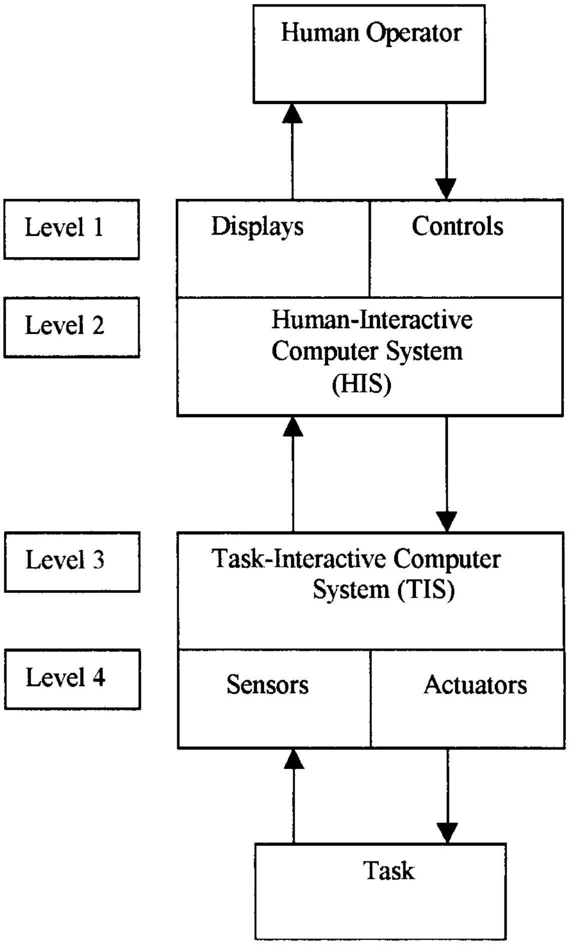


Figure 1: The basic elements of supervisory control (after Moray, 1986

Thus there is a requirement to ensure that there is appropriate mapping between psychological and physical variables at two distinct levels.

This systems model is more commonly associated with process control, and complex manufacturing systems. However, there are common elements between these domains and telepresence control because of the introduction of a computerised intermediary. Human-computer interaction provides a further intermediary in the control process, thus introducing one further level of abstraction.

These above factors have all been instrumental in influencing our approach to designing the current system. The concepts outlined above represent our conceptual framework upon which our design decisions were made.

2.2. Design of the user interface.

The objectives behind the development of the control system for RCV operations were derived from an empirical study concerned with mental workload demands of RCV operation[5]. In trials which required subjects to negotiate a series of obstacles (markers, ramps etc), using the control system provided, via a CCTV link up, and which culminated in using a telescopic boom to remove a parcel from a car seat.

Results showed that there were several problems associated with understanding the function and control system for the RCV.

- Subjects had no clear understanding of the vehicle's capabilities.
- Full utilisation of functions was rare.
- Spatial and configurational awareness about the vehicle was poor.
- Control interface was poor with no clear labelling of functions.

Subjects were also unsure as to the full capabilities and limitations of available functions. This appeared to be due to the lack of clarity of the control interface.

Many principles learned in one stage of the trial were not used again, even when the conditions demanded it.

It was clear from the results of this study that naive subjects found accurate manipulation of the vehicle during telepresence difficult. There were two principal reasons for this. Firstly users were unable to develop an accurate and fully fledged conceptual model of how the system worked. Secondly, spatial awareness, and the relative position of the vehicle in relation to surrounding objects was surprisingly poor, given that subjects were allowed to walk around the test track prior to the trials.

The results suggested that there were a number of improvements which could be made to the effective use and control of the RCV, by integrating state of the art sensor technologies with sound design principles from human-computer interface design.

Direct manipulation (DM) provides users of interactive systems with the opportunity to interact with objects much in the same way as they would interact with objects in the real world. One of the key principles of DM lies in providing users with a sense of control over the interactive world. Some of the advantages of this style of interaction are outlined below[6]

- Novices can learn functionality quickly, usually through a demonstration by a more experienced user.
- Experts can work extremely rapidly to carry out a wide range of tasks, even defining new functions and features.
- Intermittent users can retain operational concepts.
- Error messages are rarely needed.
- Users can see immediately whether their actions are furthering their goals, and if not, they can simply change the direction of their activity.
- Users have reduced anxiety because the system is comprehensible and because actions are reversible.

These principles relate to the design of desktop systems that use standardised software and are not specifically aimed at control systems for robotic operation.

3.The Control System

A control system has been developed which combines a user interface with internal and external sensory systems.

3.1. User Interface.

There are currently three levels of user interface. At the lowest level is the configuration and maintenance interface. At an intermediate level is the normal operating interface and at a higher level still is the icon-based interface.

3.1.1. Maintenance / Configuration interface.

This interface provides the operator with the option to configure the control system according to current equipment installed on the vehicle. It also allows the operator to configure the level of warning and information provided to the operator.

3.1.2. Normal operating interface.

This interface provides all the normal robot operating functions such as joint by joint and resolved motion control of the manipulator, a real time mimic showing actuator configuration, actuator joint angles, camera configuration and position and the X,Y co-ordinates of the end effector. The system also provides standard driving controls for the vehicle both via a computer keyboard and mouse and through Joysticks mounted on a hand held pendant. The user interface also provides warning of collisions of the manipulator with the structure of the vehicle. These warnings are presented both on the PC screen and as overlays on the CCTV monitor. A number of different configurations of manipulator are supported which are entered by the operator. The normal operating interface is shown in Figure 2.

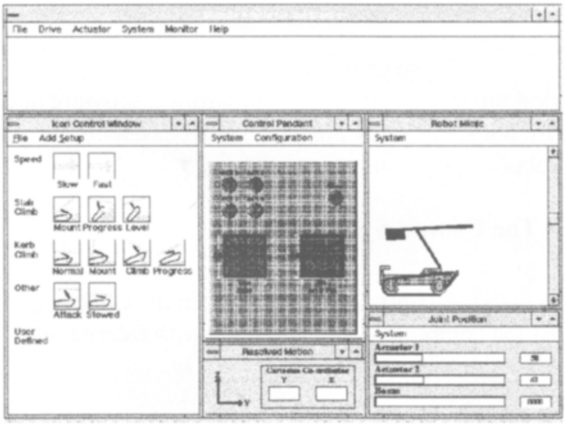

File
Drive
Actuator
System
Monitor
Help
Icon Control Window
File
Add Setup
Speed
Slow
Fast
Stair Climb
Mount
Progress
Level
Kerb Climb
Normal
Mount
Climb
Progress
Other
Attack
Stowed
User Defined
Control Pendant
System
Configuration
Robot Mimic
System
Resolved Motion
Cartesian Co-ordinates
Y
X
Y
Joint Position
System
Actuator 1
Actuator 2
Boom

Figure 2: Normal operating interface.

3.1.3. Icon interface.

In addition to the normal operating interface there is a higher level interface which allows operation of the vehicle through an icon based user interface.[7] The operator is able to control the configuration or stance of the vehicle automatically by using a series of icons linked to embedded software routines that automatically reconfigure the position and angles of the vehicles's structures. The icon interfaces shown in Figure 3.

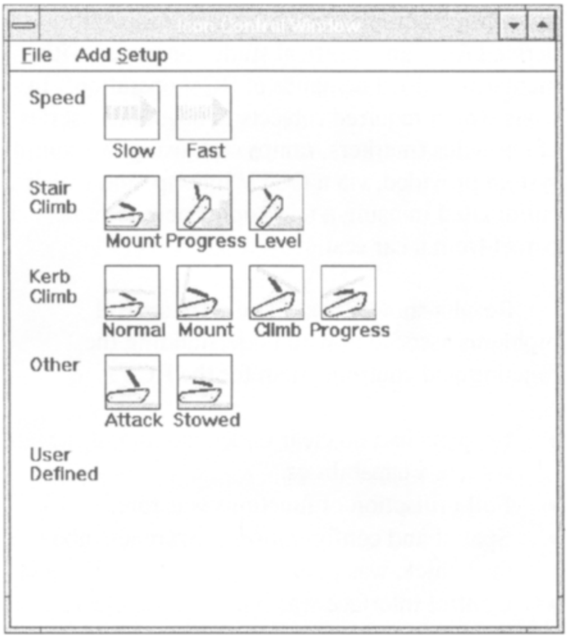

Icon Control Window
File
Add Setup
Speed
Slow
Fast
Stair Climb
Mount
Progress
Level
Kerb Climb
Normal
Mount
Climb
Progress
Other
Attack
Stowed
User Defined

Figure 3: Icon interface.

3.2. Sensory System

The sensory systems are divided into two parts. External sensors provide information about the vehicles operating environment and internal sensors provide information about the status and configuration of the vehicle.

3.2.1. External sensors

For an operator to have a true sense of the scale of the environment the vehicle is operating in some type of range measurement system is required. Ultrasonic sensors are commonly used but they have a number of disadvantages these include:

- Specular reflections due to the relatively long wavelength of sound.
- Poor angular resolution due to the limited beamwidth of available transducers.
- Propagation effects caused by atmospheric temperature gradients, humidity and air currents.
- Lack of robustness of capacitive transducers and inability of piezo-ceramic transducers to couple well with air.

Whilst the influence of many of the above problems can be reduced using advanced ultrasonic arrays[8] and signal processing[9,10]. However, the lack of robustness makes such transducers unusable in this application.

The external sensory system chosen for the control interface is an infrared scanning laser range finder manufactured by Erwin Sick of Germany. It provides a full 180° scan with a maximum range of 50m, an angular resolution of 0.5° and an accuracy of ±50mm. This data is available to the operator as a raw range map and is used by a collision avoidance system. The output for a typical scene is shown in Figure 4.

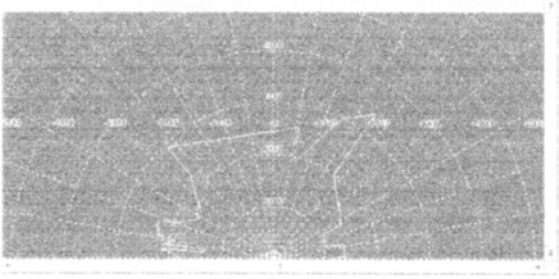


Figure 4: Output of laser scanner.

3.2.2. Internal sensors

The internal systems consist of potentiometers which provide information on the position of the cameras and the manipulator joints. A Hall-effect sensor which provides data on the amount of extension of the telescopic boom. The vehicle is also equipped with a number of limit switches which are used both for overrun protection and calibration.

Under normal operation the readings from the joint position sensors are used to drive the user interface providing updated position information ever 250mS. In addition to this the forward kinematics of the manipulator are continuously calculated for both the telescopic boom end and the motor located at the back of the telescopic boom. Both of these points are liable to collision with either the ground or the vehicle structure. If a likely collision is detected the operators control inputs are limited to those actions which will recover from the potential collision.

3.3.Collision Avoidance

The current collision avoidance system consists of a simple 'stop and warn' algorithm that provides the operator with a warning when obstacles are detected by the Laser within a preset range of the vehicle. If the operator ignores the warning then the vehicle is automatically halted if the obstacle falls within a second precept range. A much more advanced collision avoidance and navigation system has been developed which will eventually be incorporated into the demonstrator[11].

4.System Realisation

4.1. Technology Demonstrator

A technology demonstration and development vehicle has been constructed around a Wheelbarrow Mk7 Explosive Ordnance Disposal Vehicle (see figure). A number of sensors have been fitted to the vehicle to provide feed back on the position of the onboard actuators and camera along with an infra-red scanning laser range-finder mounted on the front of the vehicle. The vehicle is connected to the control station by an umbilical which caries control information, sensory information, CCTV signals, Laser power and a high speed serial data link (RS422). The data link provides control information to, and returns range data from, the Laser. The vehicle control system and user interface uses a 150 MHz Pentium computer and hand held control box. The system software is written using C operating under the windows 3.11 operating system. Sensor processing is performed by a TMS320C40 Digital Signal Processor which also generates the video overlays for the CCTV display. A block diagram of the demonstrator hardware is shown in Figure 5.

5. Conclusion

The work described in this paper provides the foundation for the development of an intuitive user interface. Future work will see the inclusion of more intelligence in the form of user support systems and the semi-autonomous navigation system currently under development[11]. Much of the system has already been demonstrated using the demonstrator vehicle and full user evaluation trials are planned.

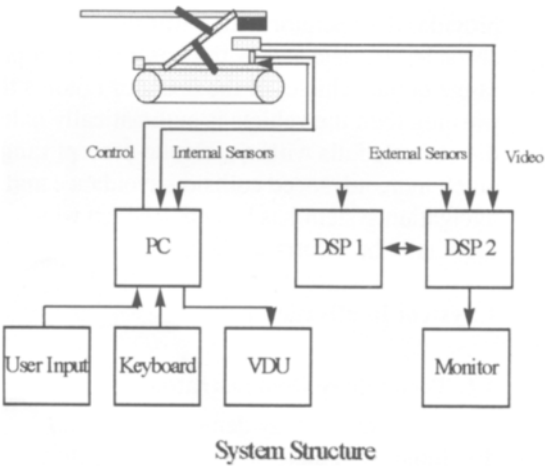


Figure 5: Technology demonstrator hardware

6. References

1. Webster, A. W. and Mistry G. C., " Robotics in the Nuclear Industry: A review of BNFL's Remote Handling and Robotics Program for the Development of Generic Techniques and Equipment", Proc MELCON 94, Malaga, August 22-25 1994, pp1479-1508.

2 . Milgram, P., Rastogi, A., and Grodski, J.J., 'Telerobotic Control Using Augmented Reality.' Proceedings 4th IEEE International Workshop on Robot and Human Communication (RO-MAN'95), 5-7 July, 1995.

3 Norman, D.A., 'Cognitive Engineering.' In 'User Centred System Design ; New Perspectives on Human-Computer Interaction,' Donald A. Norman, and Stephen W. Draper (eds). LEA, Hillsdale, NJ, London, 1986.

4. Moray, N., Monitoring behaviour and supervisory control . In K. Boff, L. Kaufman, & J. Thomas (Eds). Handbook of Perception and Human Performance (vol. 2), New York, Wiley, 1986.

5. Miles, S., The evaluation of a real time robot control interface using mental workload assessment methods. Undergraduate dissertation, Department of Manufacturing Engineering and Operations Management, University of Nottingham, 1997.

6. Shneiderman, B., 'Designing the User Interface. Strategies for Effective Human-Computer Interaction. 2nd edn, Reading, MA, Addison Wesley, 1992.

7. Brown, N. A. "The design and Implementation of an Interactive user Interface for Mobile Vehicles", Masters Thesis, Department of Manufacturing Engineering and Operations Management, University of Nottingham, 1997.

8. P. Webb, I., Gibson and C. Wykes.
Robot Guidance Using Ultrasonic Arrays,
Journal of Robotic Systems, Vol. 11, No. 8, 1994, pp681-716.

9. P. Webb and C. Wykes. High Resolution Beam Forming for Ultrasonic arrays IEEE Transactions on Robotics and Automation, Vol. 12, No. 1, 1996, pp138-145.

10. P. Webb and C. Wykes. Suppression of Second Time around Echoes in High Firing Rate Ultrasonic Transducers, NDT&E International, Vol. 28, No. 2, 1995, pp89-94.

11. C. Fayad and P. Webb. Navigation of a Semi-Autonomous Mobile Robot in a Dynamic Known Terrain, proc 29th International Symposium on Robotics, Birmingham, 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

An Intelligent Mechatronic Architecture for Hexapod Robot Control

M.J. Randall, A.G. Pipe and A.F.T. Winfield

Intelligent Autonomous Systems Engineering Laboratory, Faculty of Engineering,
University of the West of England, Coldharbour Lane, Bristol, BS16 1QY, United Kingdom

The goal of this research is to produce an autonomous hexapod, which walks over extremely irregular terrain with sparse footholds. The control architecture is hierarchical and is described from a top-down perspective. It consists of four levels: "motivation", body trajectory, kinematic planning and dynamic actuation. Vectors provide the interfaces between these levels. The architecture described is implementation-independent and thus can be applied to virtually any hexapod (or, more generally to any mobile robot). In a sense, our architecture serves as a soft computing "mechatronic interface": it provides an extra layer of control that ensures that a command sent by any electronic controller will be realised on any set of mechanical actuators so long as it is physically possible.

1. THE HEXAPOD

The goal of this research is to produce an autonomous hexapod that walks over extremely irregular terrain with sparse footholds. We have taken a biologically-inspired approach to the design, control and rough terrain strategies of the hexapod, which are loosely based on the stick insect *Carausius morosus*. To date, we have developed the hexapod's control architecture, and have obtained favourable results in simulation. Furthermore, we have constructed a hexapod robot with typical stick insect dimensions, albeit on a larger scale. This paper presents some preliminary results of the intelligent control architecture applied to a hexapod robot.

The robot has thirty degrees-of-freedom, eighteen of which are active and twelve are passive. They are comprised as follows: each of the six legs has three active degrees of freedom at the rotational joints, and one passive degree of freedom along the length of the extreme link. Further the body -centre has six degrees-of-freedom (three translational and three represented by unit vectors around each of the translational degrees of freedom). None of these six degrees of freedom is under active control but they result as a consequence of our rigid body model. The hexapod is therefore a mechanically complex system that is non-holonomic, has closed kinematic chains and built-in redundancy. The structure of one of the legs can be seen in Figure 1. Such an articulate leg design is key to achieving the accurate foot placement required of the robot.

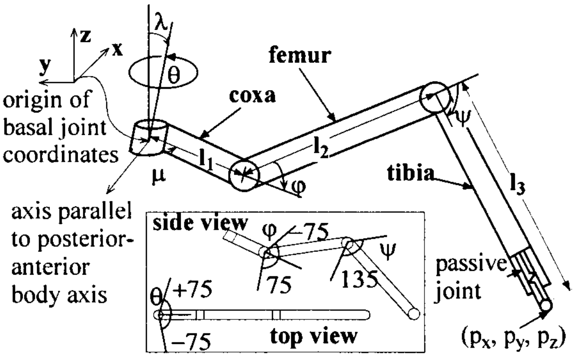


Figure 1 Leg geometry, showing joint angles (θ, ϕ, ψ), the origin of the basal joint co-ordinates frame (0, 0, 0), and the two angles describing the orientation of the leg base frame with respect to the body (λ, μ). Also shown are the foot co-ordinates (p_x, p_y, p_z) relative to the basal joint where the coxa meets the body. Inset shows joint angle limits.

The θ- and ϕ-joints are powered by standard model aircraft servos delivering a torque of 0.66 N·m, and the ψ -joint is powered with a miniature servo delivering a 0.25 N·m torque. The total leg length varies from 172mm for the middle legs to 220mm for the hind legs and these are roughly to scale.

The hexapod body consists of three segments separated by a stiffly compliant rubber spacer. Each segment (front, middle and hind) holds its respective pair of legs and is symmetric about the body's longitudinal axis. The total body length is 660mm. There are two "antennae" at the front and two "cerci" at the back of the body to detect collisions.

Similar hexapods have been described in [1-4].

2. THE CONTROL ARCHITECTURE

The control architecture is hierarchical and is described from a top-down perspective. It consists of four levels: "motivation", body trajectory, kinematic planning and dynamic actuation. Vectors provide the interfaces between these levels. From the highest level, a symbolic goal could be provided instead of a vector. The hierarchy performs a series of co-ordinate transformations: from the task-based co-ordinate system to a cognitive map, from the cognitive map to an agent-centred kinematic system, and from kinematic co-ordinates to the actuator dynamics. Each level has a return path to the level above it, which is called a "status flag". The exact nature of the status flag and sensory inputs is left open to make the architecture as generic as possible. The hierarchy can be seen in Figure 2. As such this architecture can be used to describe a whole host of control problems: from human planned or extemporaneous movements to motion planning and execution in mobile robotics.

2.1. The Motivation Level

In order to be as autonomous as possible an agent must be able to produce motivations of its own devising. Machine motivation can be expressed as an interplay between machine emotions (e.g. "fear", "curiosity") and machine needs (e.g. low power, maintenance requirement). One structure for implementing this could be a high-level behaviour selection mechanism [e.g. 5] wherein a behavioural module represents an emotion or need. To use an anthropomorphism, one could view the processes of competition (and perhaps co-operation) between behavioural modules as analogous to the pre-conscious parallel processing underlying a "stream of consciousness". The sequential selection of competition winners thereby gives rise to "conscious" motivations. An architecture such as this shares some common ground with the model for human consciousness proposed by [6].

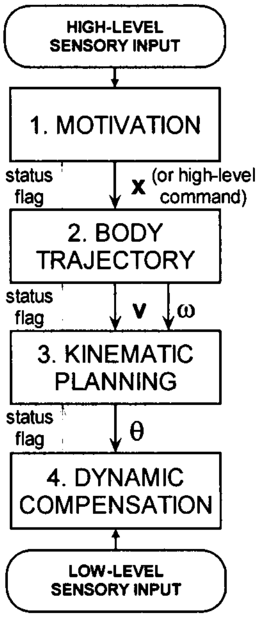


Figure 2 The Generic Control Hierarchy. High-level sensory input produces a motivation to move to say a position vector **x** for the robot. Route planning at level 2 transforms this into body-centre velocity and angular velocity vectors **v** and ω. Level 3 implements a kinematic plan producing target joint space vectors θ for the legs. Level 4 transforms these into actuator signals that minimise error between target vectors and actual position. Low level sensory information may be required to do this.

A motivation expresses the desire for attaining a current reward state. Since the context is walking machines this might be one of a set of symbolic goals to be attained such as "locating a battery charging station", "detection of structural fault in bridge", or perhaps "detection of new mine".

2.2. The Body Planning Level

Given a goal to be achieved, this level in the hierarchy must establish an appropriate body route through a partially or completely unknown environment. This navigation task has been extensively investigated in our lab [7-8].

A cognitive map is suitable for use in mobile robot navigation under the following circumstances: the agent needs to transcend basic stimulus-response activity but landmark recognition is limited to relatively crude local scene characterisation, and it is not possible for the agent to leave "markers" in its environment. These are likely to be pertinent circumstances for the applications in question here. The overall objective is to make use of the best features of two cognitive mapping forms, each having its strengths and weaknesses for different aspects of mobile robot navigation, and to combine them with the good reactive abilities, which are achievable by a behaviour-based approach.

A non-procedural metric spatial map of the agent's environment is built through interaction with it [7]. The resulting cognitive map is a continuous representation of utility for any location in terms of the likelihood of reaching positive reinforcement (reward). This is powerful because an agent could hypothesise about unvisited routes between two known locations using the global metric information contained in the map. This would not be possible using a topological representation. In the experiments presented below, mobile robots fitted with ultrasonic sensors are assumed. However, all the ideas are transferable to other sensorimotor interfaces.

The map-building architecture is based upon the adaptive heuristic critic (AHC) paradigm, which is made up of two principal components, the "action policy" and "critic". In this architecture an evolution strategy (ES) is used directly as the main exploring part of the agent, i.e. it forms the "action policy" for each movement in the maze. A radial basis function (RBF) neural network is used to store the knowledge gained from exploration and therefore forms the main part of the "critic", updated using a temporal difference (TD) learning algorithm as the search progresses. Each ES member represents a potential move from the current position, expressed in egocentrically oriented polar co-ordinates. The RBF network performs a mapping between a given continuous valued xy-coordinate and a measure of "value" or "utility" for that position.

Many experiments have been performed using this architecture [7]. One example is presented here. The start position is at the top left of Figure 3. Figure 4 shows the cognitive map that was formed. "Bad places" and obstacles appear as ridges, whilst "good" places appear as valleys. The main contributions made by this architecture are the ability to efficiently search a large action space using the ES, and to perform local generalisation about the effectively infinite state space using the RBF neural network. However globally metric mapping methods face many practical problems. Two that stand out are cumulative errors in the current position estimate and on-line computational requirements.

Perhaps the best role for a globally metric map-building architecture is in allowing an agent autonomously to acquire maps of approximate accuracy under static environmental conditions, which could then be used for off-line route planning. This involves the extraction of a procedural topological map from one of the above metric maps

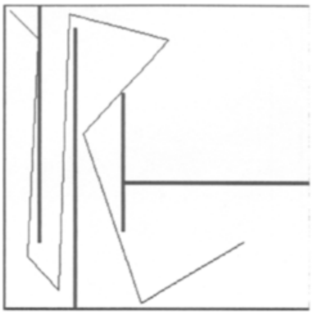


Figure 3 The procedural metric map

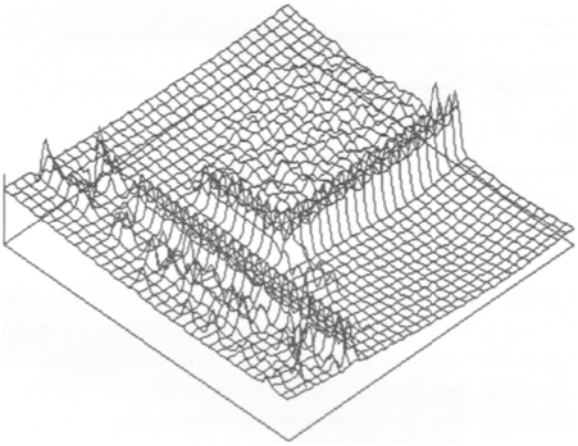


Figure 4 The example cognitive map.

and the inhibition of a lower level reactive behaviour-based system. Although it uses a weaker form of map than that described above it is computationally much less intensive and does not suffer from cumulative path integration errors. The agent should now express this trajectory in a procedural topological form, for on-line use in the presence of local environmental disturbances. At the lower level a behavioural module typically contains a small group of competencies, such as "corridor following" or "turn right at T-junction". For topological map building purposes each competency is expressed by an explicit symbolic component. A semi-ordered set of competencies may be built to form a procedural topological cognitive map. Each map represents a route through a particular maze to a goal from a given start point. In behaviour-based robotics terms, it acts as a second level behaviour selection mechanism that overrides a basic flat competitive system [9] when necessary. It inhibits behavioural modules as appropriate in order to resolve

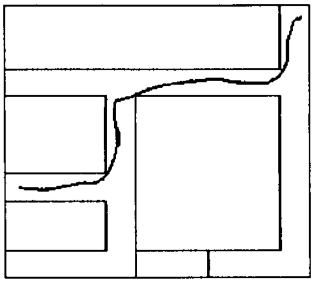


Figure 5 Maze solved using procedural topological approach

Table 1 Topological Map for multiple T-junction maze

Map category	Competency list
Default	Small-obstacle-avoidance
Non-conflict	Winding-corridor
Conflict	Right-T; Left-T; Right-T

conflicts and achieve the desired behaviour. Since it is a topological representation, it is as insensitive to local disturbances as the behavioural modules that it controls will allow. Each behavioural module contains one or more competencies (e.g. obstacle avoidance, turn left at T-junction), and is implemented using standard fuzzy logic.

Although a topological map can be completely dimensionless at a high level, it must use metric or other detailed environmental features at a lower level in order to perform place or landmark recognition. Furthermore, agents may be physically different from each other. Even if they are physically identical, they may possess different mappings between behavioural modules and the competencies contained within them. A map is therefore specified at the level of competencies and it is assumed that the agent is able to perform the translation between these and its behavioural modules. A "minimum intervention" approach is adopted in the map protocol, which leaves the basic processes of behavioural module competition undisturbed unless it is necessary to resolve a conflict. A map specifies the set of non-conflicting competencies required to complete a route through a maze, plus an *ordered* list of competence *inhibitions* (in the sense of [5]) for conflicting modules. "Place identifiers" trigger progress through this list: rules embedded in the fuzzy logic of behavioural modules. A map consists of three main parts specifying the competencies required to complete a route through a maze as seen in Figure 5.

The competencies were acquired autonomously by the agent, using the procedure described in [9]. Again, one selected result from more thorough testing is presented. To evaluate the architecture proposed above a 3m×3m multiple T-junction maze was devised. To proceed successfully from the start at the left of the middle corridor to the goal at the top right corner the agent must turn left at the first junction, right at the next and left at the last junction (See Table 1). A metric map of the complete maze was created using the non-procedural map-builder. This was then used in the processes of extracting behavioural modules for turning right at a T-junction, turning left at a T-junction and following a winding corridor, establishing appropriate place recognizer rules, and building competency lists for the non-conflict and conflict categories of the topological map shown in Figure 1. Note that the conflict category is an ordered list consisting of clauses separated by ";" which list competencies to be inhibited. For T-junctions, there will only be one competency to suppress, but more generally, there could be more than one. The maze and body trajectory generated by the agent (left to right) is shown in Figure 5. By contrast to the non-procedural metric map-builder, this architecture is quite robust to local environmental dynamics [7]. So far, this research has been confined to 2D worlds but some initial work has been undertaken to expand it to the 3D environment of a typical walking robot.

2.3. Kinematic Planning: Gait Control

The body route level of the hierarchy transforms the high-level goal of the system into a velocity and angular velocity vector. This information is used to determine the speed of the hexapod. For a walking system, body speed is expressed in terms of gait. Essentially the gait determines the timing of leg swings in relation to each other (co-ordination) but gives no information as to leg placement or trajectory. Hence, kinematic planning is decomposed into gait control and leg trajectory generation.

The gait controller is based on a very simple biological model proposed in [10]. There are other more complicated (and thus more effective) models which we have also investigated [11]. However, simplicity is a great advantage in a highly rough terrain environment since most of the computational

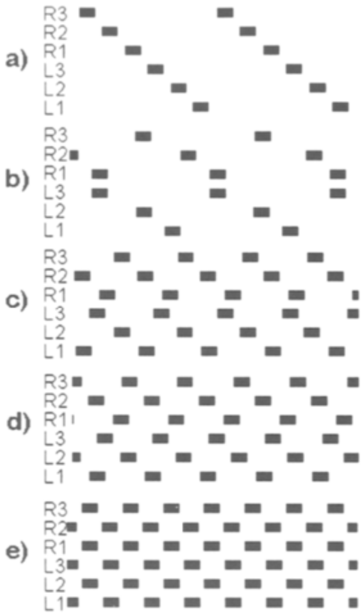


Figure 6 Gait Patterns generated from the Graham model. (a) Wave gait with no coupling. As the coupling is increased the left and right metachronal waves overlap more with a continuum of patterns ranging through (b) to (e). (c) and (d) are tetrapod gaits and (e) is the typical tripod gait. Black bars represent leg swings.

effort required is to achieve accurate foot placement (leg trajectory control). The model is effective enough to produce the range of gaits observed in insects by Wilson [12], which guarantees static stability—clearly an important feature of the locomotion.

2.4. Kinematic Planning: Leg Trajectory Control

The trajectory planner is derived from studies of human arm trajectories. There are two reasons for choosing such a starting point. First, it is intended that the artificial hexapod reduce "wear and tear" on its mechanical components by employing some appropriate cost-minimising function observed in biological systems. Second, it is intended to generate trajectories that appear "natural", as a human observer of the robot would perceive it. Also, the studies provide many insights into some human cognitive level organisational principles behind trajectory determination. In fact, the "aesthetic" features of natural movements have been shown to correspond to just such organisational principles or rules [13].

In this work, leg trajectory planning is accomplished in kinematic terms. A large body of evidence suggests this is the correct approach. The full justification for this position and the choice of optimisation policy described below is given in [8]. In order to generate a unique protraction trajectory, we have chosen to minimise the mean-squared jerk over the total duration of three-dimensional movement. This is an extension of the work of [14] where "jerk" is defined as the time derivative of acceleration. It turns out that the predicted trajectories are very computationally expensive. Therefore, we have introduced a number of simplifications that produce up to 6% more "jerky" trajectories but still provide accurate foot placement and are much more computationally simple [8].

2.5. Dynamic Compensation: Adaptive Neural Control

Finally, at the level of each joint, we plan to implement adaptive neural controllers to compensate for errors in the planned leg trajectories. Thus, this level transforms the planned trajectories into a series of actuator signals that compensate for errors in target values and actual signals. The theory of adaptive neural control and some preliminary results showing how we have implemented it in this architecture are given in the companion paper [15]. In summary, the adaptive neural controller will learn on-line and in a stable manner to compensate for errors in foot position and velocity. By monitoring torques, the adaptive neural controller will deal locally with certain aspects of terrain roughness, thus emulating some of the rough terrain strategies observed in insects [cf. 1]. The effect of this layer will allow the architecture to be described as a soft computing "mechatronic interface". This is because it provides an extra layer of control that ensures that a command sent by any electronic controller will be realised on any set of mechanical actuators so long as it is physically possible.

2.6. Implementation

This control architecture is implemented on a Motorola 68332-based board designed in the Faculty of Engineering at the University of the West of England, Bristol This on-board computer is clocked at a rate of 20MHz. The controller possesses an on-chip Time Processor Unit (TPU) which allows for conversion between integer values representing joint positions and a PPM signal to be sent to a servo-motor input. The controller also possesses a Queued Serial Module (QSM) which provides for synchronous (SPI) and asynchronous (RS232) communications. The former is used for communication with some on-board peripheral devices, whilst the latter is used for transmitting the results of experiments to a host computer. The board is equipped with A/D converters, which are connected over the SPI line, for measuring the servo-motor position feedback signals. Software has been developed in the C language, using a PC-resident Microtec Research "XRAY" cross-development suite.

3. DISCUSSION AND APPLICATIONS

Distributed control for hexapod robots has been used since the early 1980s [16]. The popularity of neural networks since then means that they have been employed by several authors as a control technique [e.g. 2-4]. The nature of the implementation differs here since we have used RBF neural networks. Steuer et al. [2] describe a similar hierarchy but without the first level, which enables the robot to achieve a greater level of autonomy. There are at least three degrees of autonomy that are of interest to researchers in walking or mobile robotics. The first (and lowest) is that of "power" autonomy, where the robot is a self-contained unit that does not require an umbilicus providing power.

The second degree of autonomy might be described as "intelligence" or (perhaps less contentiously as "adaptability") in that the agent is able to adapt to a potentially dynamic environment and achieve a goal. This is our aim and we believe this is achieved by our architecture using soft computing techniques.

The third (and highest) degree of autonomy might be described as "independence" where the agent is able to repair itself from significant (though not fatal) damage and does not require maintenance from an external source. This third level of autonomy is achieved by insects (in fact all animals).

Once power and intelligence autonomy are achieved (although power autonomy may not necessarily come first), several applications are opened up that seem elusive otherwise. These include exploration (as in Antarctica, the seabed, or on a planetary surface); humanitarian de-mining; and industrial inspection and maintenance of for instance nuclear utilities, inside ductwork and pipes, the underside of oilrigs or large ships.

REFERENCES

1. K.S. Espenschied, R.D. Quinn, R.D. Beer, H.J. Chiel, Robs and Auton Sys, 18, 59-64, (1996)
2. J. Steuer, F. Pfeiffer, Proc Euromech 375, (1998) 141-148.
3. M. Frik, M. Guddat, D. Losch, M. Karataş, Proc Euromech 375, (1998) 108-115.
4. K. Berns, R. Dillman, S. Piekenbrock, Robotics and Autonomous Systems, 14, 233-244. (1995)
5. R.A. Brooks, IEEE J. Robotics and Automation. RA-2(1), (1986).
6. D.C. Dennett, Consciousness Explained, Penguin, 1993.
7. A.G. Pipe, PhD thesis, 'Reinforcement Learning and Knowledge Transformation in Mobile Robotics', UWE, Bristol, UK., 1997
8. M.J. Randall, A.G. Pipe, Proc. Intl Workshop Adv Robotics and Intelligent Machines 1997.
9. A.G. Pipe, A. Winfield, From Animals to Animats 4, MIT Press, (1996) 233-242
10. D. Graham, J Comp Physiol. 81, (1977) 23-52
11. M.J. Randall, A.G. Pipe, A. Md. Gh. Muttalib, A.F.T. Winfield, Proc Euromech 375, (1998) 124-132.
12. D.M. Wilson, Annual Review of Entomology 11, 103-122 (1966)
13. P. Viviani, in Attention and Performance XIV MIT Press, 1992.
14. T. Flash, N. Hogan, .J Neuroscience, 5(7) 1688-1703 (1985).
15. M. J. Randall, A.G. Pipe, A.F.T. Winfield, 'Adaptive Neural Control of Hexapod Leg Trajectories' (same volume).
16. M. D. Donner, Real-Time Control of Walking, Birkhäuser Boston, 1987.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Behavior-based Control of a Four Legged Walking Robot

L. Pettersson[a], K. Jansson[a], H. Rehbinder[b], and J. Wikander[a]

[a] Department of Machine Design, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

[b] Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

The paper describes a behavior-based control system architecture for the four legged walking robot Sleipner III. We investigate how the behavior-based approach can be used for control and gait generation of a quadruped walking in uneven terrain. An architecture with behaviors distributed on two levels - leg level and body level are described. On leg level the legs are to a large extent seen as independent agents in order to minimize central coordination and control. Results from simulations and data on the robot are given.

1. INTRODUCTION

Walking using a purely reactive controller was shown to work by Brooks (1989) who implemented the behaviors as a distributed network of finite state machines. Beer et al (1993) used behavior-based approach for control of a biologically inspired robot by an artificial neural network imitating the cockroach nervous system. Other examples where a behavior based approach has been used to generate walking patterns include Wettergreen et al (1995) and Luk et al (1996). All these examples of behavior-based control and gait generation are for six or eight legged robots. Gait generation for quadrupeds by some central leg sequencing mechanism, using an algorithmic rather than behavior-based approach, is well studied [Song & Chen (1991), de Santos & Jimenez (1995), Hirose & Kunieda (1991)]. Very little, however, has been done in studying how reactive behaviors can be used to generate gaits and control walking of quadrupeds.

A gait that follows a predefined sequencing of the legs is called a *fixed* gait in contrast to a *free* gait without any predefined leg sequencing imposed. For two reasons we will avoid central coordination or sequencing of the legs and instead generate walking patterns by distributed reflex-like behaviors. Firstly it is essential to reduce communication between legs and body of the robot in order to minimize the need for cabling and real-time problems due to excessive message sending. Secondly in difficult terrain the robot cannot be expected to use a fixed gait for any extended period of time. Viewing the four legs as much as possible as individual agents will in a natural way imply generation of a free gait.

When the robot is walking in regular terrain a free gait will emerge if the behaviors are appropriately chosen but the goal should never be to generate some predefined gait but to achieve robust and swift locomotion in the terrain. This type of behavior-based control system will be implementing control levels between low level control of individual actuators and high level control like path planning, foraging or following behaviors.

2. SYSTEM DESCRIPTION

The robot Sleipner III that is currently being built will be used as a platform for research on swift locomotion in difficult terrain. A requirement is that the robot is fully autonomous with all processing as well as power supply onboard. The quadruped is designed to be able to walk at a speed of approximately 2 m/s.

This makes is suitable for studying both statically stable and dynamically stable walking.

The total weight of the robot is 62 kg. The motors are placed so that the weight and inertia of the legs are minimized. The robot is 80 cm high, 80 cm in length, and have legs of total length 60 cm.

The system is modular and distributed in its structure and described by division into two architectural levels - body level and leg level. Behaviors are as much as possible implemented at the leg level and executed in processors physically located on the legs to minimize communication and cabling from a central computer to the legs. A distributed structure with the four legs being independent agents are thus strived for.

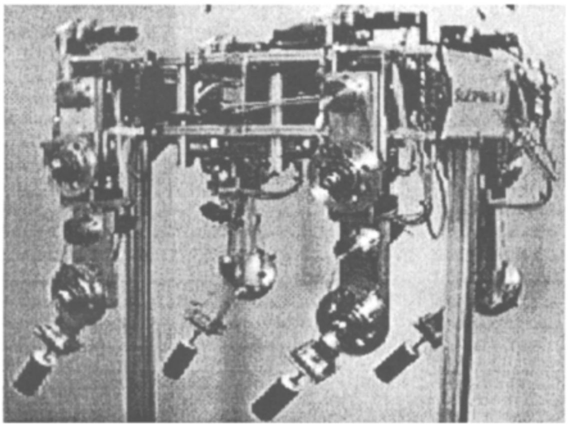


Figure 1. Sleipner III

2.1. Actuators

The four feet are individually controllable with three DOF. The configuration of each leg is described by three angles (hip flexion, knee flexion and hip abduction). The feet have point ground contact.

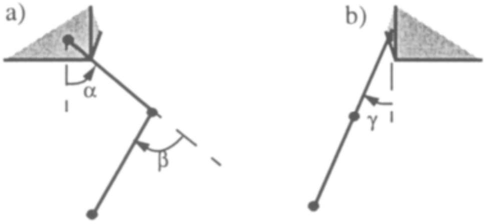


Figure 2. Configuration of one leg, a) side view, b) front view, with hip flexion angle (α), knee flexion angle (β) and hip abduction angle (γ)

The leg level behaviors are using a robot fixed cartesian coordinate system. It is straight forward to calculate the Jacobian and thereby getting the difference in leg angles (α, β, γ) needed to get a certain movement of the foot position relative to the body in cartesian coordinates. The three leg motors, each equipped with an individual servo, will be actuated accordingly.

The aim of the body level behaviors are to move the robot body in the world. This is achieved by calculating the movement of each foot relative to the body needed to achieve a certain movement of the robot body and actuating the motors of the four legs accordingly.

2.2. Sensors

The robot has no sensors for range sensing like vision or laser scanning but the robot has a number of other sensors. High level sensors will be added later but for now the robot is a “blind walker”.

A number of leg level sensors are used by the behaviors:

FootPosition: Position sensor in each joint by which the configuration of the leg and position of the foot relative to the body is known. There are encoders mounted both on the motor axes and the joints. With a resolution of 1024 pulses per revolution and by using all flanks of the pulses a resolution of 4096 steps/rev of joint angles are obtained.

GroundContact: Contact sensor in the foot that indicates if the foot has ground contact.

ObstacleHit: Contact sensor at front of the leg that indicates when the leg hits an obstacle.

LegWeight: There are force sensors in the feet used to measure the supporting weight on the leg.The force sensors have a resolution of 1024 steps in the interval 0 to 2200 N.

The following body level sensor information is used by the behaviors:

Orientation: Inclinometer that measures the orientation of the robot body relative to the vertical line (pitch and roll angles).

A rate gyro and accelerometer that give the acceleration and velocity of the robot on the body level will be included in the near future but are used by the simple gait-generating behaviors described here.

3. BEHAVIORAL ARCHITECTURE

The architectural design is inspired by Brooks experimentally driven design procedure used during the development of Genghis [Brooks (1989)]. We developed our system in an incremental manner, starting with very simple tasks like "walk along a straight line" and successively added new competencies as work progressed.

3.1. Classes of Behaviors

Behaviors are running on either of two levels - leg level or body level.

Leg level behaviors are run with one instance for each leg independent of the instances running for the other legs. Each leg is thus controlled individually assuring minimal coordination and communication between the legs.

Body level behaviors are controlling the robot body by controlling the four legs simultaneously. The body level behaviors does not need any knowledge of the behaviors running at the leg level in the controller of the individual legs.

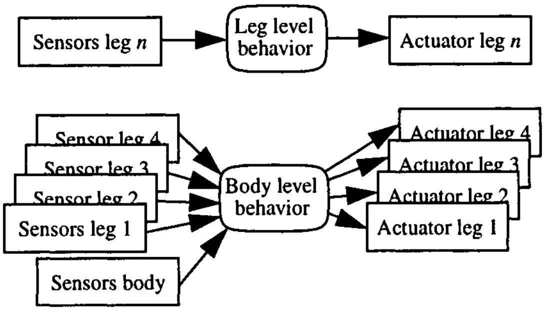


Figure 3. Sensing - actuation coupling for leg and body level behaviors respectively

The behavioral architecture includes three classes of behaviors - *step behaviors*, *posture behaviors* and *navigation behaviors*. Step behaviors are run on leg level while posture behaviors and navigation behaviors are run on body level. The behaviors included are described in the following sections and an overview presented in Table 1 on page 5.

Examples of tasks accomplished by step behaviors:

- Swinging legs forward
- Putting feet down
- Retracting feet when hitting a obstacles

Posture behaviors accomplish for example:

- Assuring stability
- Keeping a certain posture of the body
- Assuring clearance by keeping the body at a certain height above the ground

Navigation behaviors accomplish:

- Turning the robot toward some goal
- Avoiding obstacles

3.2. Behavior Coordination

All behaviors are running as separate processes with no communication between behaviors. The behaviors communicate only by their way of affecting the configuration of the robot. The leg level behaviors are supposed to execute in processes running on distributed processors located in each leg.

The behaviors are all independently trying to control the actuators and the total control signal will be generated by coordinating individual control signals. The behavior coordination mechanism, or arbitration, is a mix between a subsumption architecture and a schema architecture in the sense that the behaviors are coordinated partly via suppression and partly via summation. Between potentially conflicting behaviors there is a fixed suppression relation while other behaviors may simultaneously actuate a certain motor. The total control signal is then the sum of the individual behavioral control signals.

There is no explicit programming of leg sequencing or leg movement for walking in different situations. The gait or sequencing of the legs is instead emerging from the individual leg and body level behaviors competing to move the legs.

3.3. Step Behaviors

The step behaviors included are all trying to control the leg in a potentially conflicting way and suppression relations are introduced.

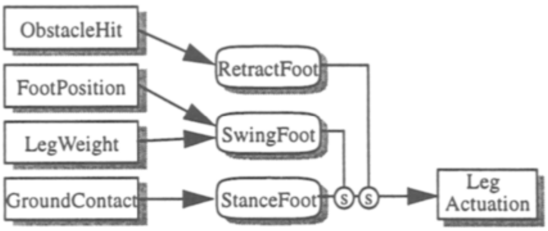


Figure 4. Suppression relations among the step behaviors

StanceFoot: This behavior keeps the robot standing. The foot is kept on the ground when it has ground contact and is lowered vertically otherwise.

SwingFoot: This behavior suppresses the StanceFoot behavior and swings the foot following a predefined trajectory. The swing trajectory is defined by the foot making a vertical lift-off and then moving to a specific point relative to the body defined by step height and step length. It should be noted that the predefined trajectory of this behavior is only one component of the final swing motion of the foot. The foot motion will also be affected by other behaviors.

During normal forward walk the step length is greater that zero. Zero step length means walking on the spot. Under certain circumstances negative or zero step length is useful, e.g. when the foot has not been able to swing forward for a certain time an insurmountable obstacle is assumed and the step length is set to zero so that the robot can turn away from the obstacle.

The SwingFoot behavior is activated when a foot has ground contact without being a supporting foot, i.e. with no dynamic weight on the leg so that the leg can be lifted without the robot falling over. Two restrictions have been used in order to avoid unnecessary swinging of the legs. First the swing behavior will only be activated if the foot is currently located far enough back so that the step length will exceed a certain limit. Secondly a short delay will inhibit activation after one swing phase is completed.

RetractFoot: This behavior retracts the foot as a reflex when an obstacle is detected by the contact sensor at the front of the leg.

3.4. Posture Behaviors

None of the posture behaviors included here are conflicting since they all affect different aspects of the body position. Thus we will not need any suppressing relation between them.

MoveBody: Moves the robot body by moving the four feet simultaneously and equally with respect to the robot body. The effect is that the robot body moves while the position of the feet in the terrain are unchanged. The behavior positions the center of gravity (COG) of the robot at an appropriate position relative to the feet. The COG will be kept inside the triangle given by two of the legs and the point of intersection of the two lines given by the diagonal legs. The stability margin can vary between 0 and 1 with the highest value denoting the center of gravity located in the middle of this triangle while the lowest value denoting the center of gravity located at the intersection of the diagonals.

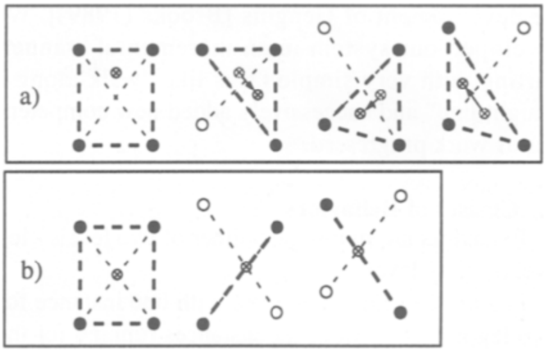


Figure 5. Positioning of the center of gravity relative to the four feet for: a) statically stable walking, and b) semistable walking.

Figure 5a illustrates the result of stability margin grater that zero. If the triangle given by the two front legs and the intersection of the diagonals is used the two front legs will be supporting legs and thus inhibited from swinging. Only one of the rear legs will be able to swing. By moving the COG a strict sequencing of the legs can be achieved and a creeping gait is generated without explicit communication between behaviors.

Figure 5b illustrates free gait generation by using zero stability margin. This can make semistable gaits for dynamic walking possible. With the COG at the intersection of the diagonals a trotting gait will be generated on planar ground with the diagonal legs swinging synchronously.

KeepHeight: Keeps the robot body at a suitable height above the ground. The body height relative to the ground is estimated using the vertical position of the feet currently having ground contact.

KeepLevel: Controls the robot pitch and roll angles by turning the robot body around the COG by coordinated lowering/raising of the feet. Appropriate pitch and roll angles are calculated using the orientation sensors together with an estimate of the slope of the ground based on the current position of the feet. The pitch angle is typically chosen so that the body follows the slope of the ground (pitch factor 1), while the body is kept horizontal with respect to rotation around the longitudinal axis (roll factor 0).

Behavior	Class of Behavior	Important Parameters	Suppressed Behaviors	Activation (Sensor)	Affected DOF (Actuation)
StanceFoot	Step	Movement velocity		GroundContact	Foot Position (vertical)
SwingFoot	Step	Movement velocity Activation delay Step length Step height	StanceFoot	FootPosition LegWeight	Foot Position
RetractFoot	Step	Retraction velocity Retraction direction Retraction distance	SwingFoot StanceFoot	ObstacleHit	Foot Position
MoveBody	Posture	Stability Margin		FootPosition Orientation	Body Position (horizontal)
KeepHeight	Posture	Raising Velocity Height		FootPosition (vertical)	Body Position (vertical)
KeepLevel	Posture	Turning Velocity Pitch Factor Roll Factor		FootPosition Orientation	Body Angle (pitch,roll)
TurnToGoal	Navigation	Turning Velocity			Body Angle (direction)
TurnFromObstacle	Navigation	Turning Velocity Activation Delay Activation Time	TurnToGoal		Body Angle (direction)

Table 1. Overview of behaviors included in the architecture

3.5. Navigation Behaviors

While the posture behaviors are restricted to body internal coordinates the navigation behaviors use some knowledge about the robots position in world coordinates.

TurnToGoal: Turns the body of the robot around the centre of gravity towards a specified goal.

TurnFromObstacle: This behavior is activated when the robot has not been able to swing its feet and walk forward after a number of trials. Then an insurmountable obstacle is assumed and the robot will turn its body away from the obstacle. If the obstacle was detected by the left foot the robot is turned right and vice versa. The behavior will be deactivated as soon as a normal foothold is detected

4. DISCUSSION

We have not included any behaviors to cope with aspects of the ground such as soft ground, slippery surface or limited supporting capacity. Our approach is to have the robot walk safely over rough but firm ground and then to introduce new competencies by adding behaviors in an incremental manner.

Accurate calculation of the dynamics of the robot have been sacrificed in order to gain computational speed. The idea is that by controlling the robot by simple distributed behaviors and by having high speed actuators in combination with low weight legs falling of the robot can be avoided by quick enough leg motions. Dynamic stability could thereby be achieved.

A problem is that the behaviors have to be appropriately tuned and adapted to the terrain. Further work on the control system architecture includes implementation of some learning mechanism for this as well as utilization of more sophisticated sensors and improvement of robustness to sensor errors.

5. SIMULATION RESULTS

We have been using a simulated model of the robot and the terrain. In simulation the robot configuration is specified by the robot position in the world

(6 DOF) and each leg is specified by the angles of its three joints according to the kinematics described in section 2.1. The weight and inertia of the legs have been assumed to be zero as the legs are designed to minimize the weight of the legs below the hip.

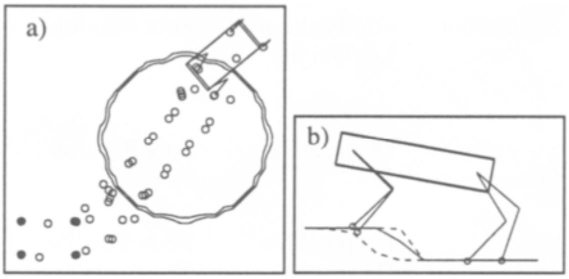


Figure 6. Robot walking over a low plateau towards a goal in the upper right corner of the figure. a) Path generated by the TurnToGoal behavior (top view). b) KeepLevel and KeepHeight behaviors maintaining an appropriate posture (side view).

The terrain is described by a Digital Elevation Map (DEM) of the terrain. Perfect sensory data, according to section 2.2, is assumed in the simulations. The sensory input are calculated based on the configuration of the robot in the terrain.

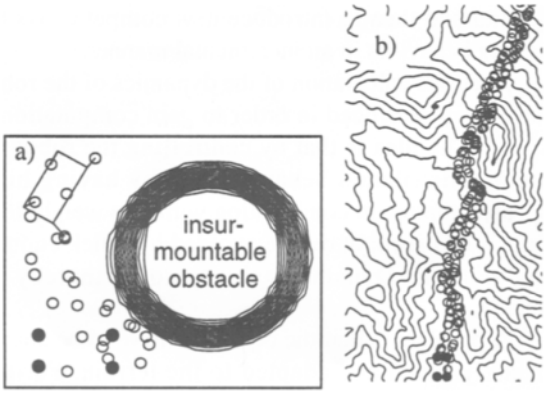


Figure 7. Robot walking towards a goal in the upper right corner of the figure. a) TurnToGoal together with TurnFromObstacle behavior generating a path towards the goal (top view). b) Path generated in a DEM taken from natural terrain.

Implementation of the control system has been done in Matlab and C. The Envision environment has also been used for simulation and animation.

6. CONCLUSION

The behavior-based approach [Arkin (1998)] is a promising strategy for autonomous robot control. We have described a possible reactive architecture and shown by simulation that it can be used for control of quadruped walking. We believe that it will prove to be a successful strategy for control and gait generation for Sleipner III. Further studies will investigate if it will be possible to generate swift dynamic walk in difficult terrain.

7. ACKNOWLEDGEMENTS

This research has been done within Centre for Autonomous Systems (CAS) which is sponsored by the Strategic Research Foundation.

REFERENCES

1. Arkin R.C., *Behavior-Based Robotics,* ISBN 0-262-01165-4, MIT Press 1998.
2. Beer R.D., Ritzmann R.E., McKenna T., *Biological Neural Networks in Invertebrate Neuroethology and Robotics*, Academic Press 1993.
3. Brooks R., A robot that walks: Emergent behavior from a Carefully Evolved Network, *Neural Computation 1:2*, 1989.
4. Hirose S., Kunieda O., Generalized Standard Foot Trajectory for a Quadruped Walking Vehicle, *The Int. Journal of Robotics Research*, 1991
5. Luk B.L., Collie A.A., Piéfort V., Virk G.S., Robug III: A Tele-operated Climbing and Walking Robot, *UKACC Int. Conf. on Control*, 1996.
6. de Santos P.G., Jimenez M.A., Generation of Discontinous Gaits for Quadruped Walking Vehicles, *Journal of Robotic Systems 12*, 1995
7. Song S-M., Chen Y-D., A Free Gait Algorithm for Quadrupedal Walking Machines, *Journal of Terramechanics*, 1991
8. Wettergreen D., Pangels H., and Bares J., Behavior-based Gait Execution for the Dante II Walking Robot, *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Biologically inspired design and construction of a compliant robot leg for dynamic walking

F. Hardarson[a], K. Benjelloun[a], T. Wadden[b] and J. Wikander[a]

[a]Mechatronics division, Dept. of Machine Design, Royal Institute of Technology, Stockholm, Sweden

[b]SANS, Dept. of Numerical Analysis and Computing Science,Royal Institute of Technology, Stockholm, Sweden

This paper describes the mechanical design of M_3L, a single leg prototype for a dynamically stable four legged robot. A discussion on how biological inspiration has been used in the design of the mechanics, based on observations of fast running mammals, is included. The M_3L resembles a mammal leg, such as cat's or dog's, in configuration and kinematics. It is articulated with three degrees of freedom, two at the hip and one at the knee. Muscle like properties are provided by DC-motors in series with torsional rubber springs which smooth out impact forces during running and have a potential for storing energy thereby increasing the efficiency. Weight and inertia are kept as low as possible to increase energy efficiency and reduce inertial forces on the body. The total length and weight of M_3L is 67 cm and 6.8 kg respectively. Experimental evaluation of M_3L's characteristics is included.

1. INTRODUCTION

Two main reasons can be found for research on legged locomotion. First there is an interest of building mobile robots with greater mobility than the more conventional robots that use wheels or tracks [5][9][11][13][14][16]. The ability of legged robots to adapt to uneven terrain by using isolated footholds while providing active suspension gives them an advantage in tranversability and smoothness of ride. They also have less destructive effect on the ground due to the fact that the locomotion principle doesn't involve slipping (at least not intentional) like wheels or tracks do when turning. This is an advantage when e.g. travelling in sensitive forest terrain or on a new wool carpet. The second reason is to understand the principles that animals use for walking such as the mechanics of the muscle-skeleton system and the neural system controlling it [1][3][12][15]. These two groups have learnt a lot from each other, the first in terms of how to design their systems and the second in terms of being able to test their hypotheses on real systems.

This paper will describe the mechanical design of M_3L, a single prototype leg for a four legged walking robot that will be capable of moving in difficult terrain and walk or run dynamically, and some issues relating to its control. The role of compliance in walking and especially running will be discussed and some tests with the compliance of the robot leg presented. The paper belongs to the first category, i.e. building robots with better mobility, but it will also discuss some valuable results from biology and try to show how they are relevant to the design of the mechanical system and the control system.

The functionality of the leg can be described in simple terms as to propel the body forward while supporting its weight when in ground contact and swing the leg forward and reposition it for the next step after the support phase is over. The mechanical structure of the leg has to be able to provide the robot with the necessary range of motion, speed and force while being efficient and robust against shocks and accidents. The control system has to provide smooth and efficient motion of the leg while being robust against disturbances such as ground variation and changes in the weight the leg has to support. Combined design issues for the mechanics and the controller are dealing with impact forces that the leg suffers at each step and unexpected collisions with obstacles. There exists a trade-off between how much should be accomplished by the mechanical system and how much by the controller

The paper is organized as follows: section 2 describes some relevant issues from studies of biological systems, section 3 describes the mechanical design of M_3L, section 4 presents experimental results and finally section 5 draws conclusions and describes future work.

2. DESIGN CONSIDERATIONS AND BIOLOGICAL INSPIRATION

There is a vast number of legged systems that can walk and run successfully in all sorts of terrain from marshes to mountain peaks. These are of course biological systems and although it is not necessary to try to imitate them in detail, in fact nearly impossible with current technology and understanding, they provide useful solutions and inspiration for the design of legged robots.

2.1. Mechanics

It is natural to look at mammals such as cats and dogs for the basic configuration as the leg is intended for a four legged robot. Most walking robots that have been built have a sprawled leg configuration similar to insects where the feet are placed wide off the body which is ideal for stability but less efficient for running as the legs usually move in an arc in the horizontal plane around the hip joint. The legs of fast running mammals swing in the vertical plane and have the feet underneath the body close to projected centerline of the body on the ground which reduces the tipping torques when only one or two legs are in support during dynamic walking or running.

As the legs' main motion is in the vertical plane, they are able to benefit from their passive properties. It has been shown that some anthropomorphic legged mechanisms can exhibit stable walk on an inclined slope without any control or actuation apart from gravity [4]. This suggests that the mechanical parameters such as length of links and mass could influence the motion and stepping frequency of the leg. This agrees with observations from nature, animals don't vary their step frequency so much during walking or running, instead they vary their stride length and the length of the support phase [15]. By bending the knee and thereby changing the moment of inertia around the hip joint they can vary the legs' pendulum frequency during the swing phase. Although an actuated system is not able to swing freely due to friction in actuators and gears, unless there is some possibility of coupling the joints free, it provides an insight in how motion can be achieved in an efficient manner.

The distribution of muscles on the legs is also noticeable. The strongest and thereby the largest muscles are situated high up and as close to the body as possible and then tendons are used to connect the muscles to the joints they work around. The reason is obvious from an efficiency point of view as this reduces the moment of inertia and the inertial forces of the swinging leg on the body. Placing actuators out on the joints should therefore be avoided and low weight of the moving parts should be preferred from an efficiency point of view and for stability.

2.2. Compliance

The role of compliance in walking and especially running has been shown to be important as it enables the animals to store energy, travel smoothly and provides them with shock tolerance.

The compliance comes from the elastic properties of the muscle-tendon system in which, in a simplified view, the muscle can be seen as the actuator and the tendon as a spring working in series with the actuator although the muscle itself has some elastic properties where its stiffness can be varied by neural stimulus. As muscles can only contract and not extend they have to be arranged paired on each side of a joint. By coactivating antagonist (opposing) musclegroups around a joint the torque and stiffness around the joint can be varied simultaneously [6][15]. The torque T around a joint, where linear relationship between neural activity and force in each muscle is assumed, can be written as [6]

$$T = T_{max}(M_f - M_e) - K(M_f + M_e)q \qquad (1)$$

where T_{max} is the maximum available muscle torque, K is the stiffness, q is the joint angle and M_e and M_f are the neural activity in the antagonist muscle groups, extensor and flexor respectively scaled to be between 0 and 1. The torque can then be controlled by setting $(M_f - M_e)$ and the stiffness by setting $(M_f + M_e)$. This model allows for an adaptive scheme to control torque and stiffness as suggested by Hogan [6] and has been used by Wadden and Ekeberg with an biologically inspired neural controller [15].

Alexander discusses three uses of springs in legged locomotion where the first two examples are based on his observations on how animals use the elastic properties of their muscle-tendon system [1].

The first example Alexander discusses is the similarity of a leg of an animal to a pogo stick as at landing the leg complies and stores potential energy in its muscles and tendons which again is released at take-off. Secondly Alexander discusses the use of return springs in the joints that are used when the legs have to change the direction of motion at the end or beginning of a step.

The third example Alexander mentions is the use of a compliant foot pad to reduce the impact of the foot with the ground. An enlightening example is how the hoof of a horse damps and absorbs shocks. At impact the lowest joint on the leg (just above the hoof) will move from 45 degrees to almost horizontal position to dampen the impact while the hoof itself deforms by widening at the back thereby absorbing energy from the shock. Related to this is an observation from control of manipulators that for stable interaction of the endpoint with an environment some initial compliance has to be present in the system [8].

3. MECHANICAL DESIGN

This chapter will describe the mechanical design of the *M₃L*. It differs from most other robot legs in that it is designed to utilize elastic actuation and provide dynamic stability. Previously, elastic actuation has been used by Raibert's hopping robots [13] and by Pratt in the spring turkey and flamingo robots [10][11]. Yamaguchi and Takanishi developed antagonistic joints by using two motors and two nonlinear springs on each joint of a biped robot [16]. A similar leg as *M₃L* has been built by Mennitto and Buehler who used CLADD transmission system which converts rotational to translational motion, combined with cables and pulleys [9]. A more detailed description of the mechanics of *M₃L* can be found in [2].

3.1. Basic characteristics

The leg has a configuration similar to a mammal leg as it is articulated and the kinematics are similar to fast running mammals such as cats or dogs, see figure 1. It has three rotational degrees of freedom where the first joint (seen from the body) is an adduction/abduction joint at the hip i.e. motion of the foot sideways, the second is an extension/flexion joint also at the hip i.e. motion in the forward/backward direction, and the third joint is an extension/flexion joint at the knee. These three joints will be referred to as the abduction joint, hip joint and knee joint. This configuration decouples the degrees of freedom where the hip and knee joint provide the propulsion and the abduction is for balance and turning of the robot. This arrangement makes it possible to take advantage of the leg's passive properties as discussed in the previous section.

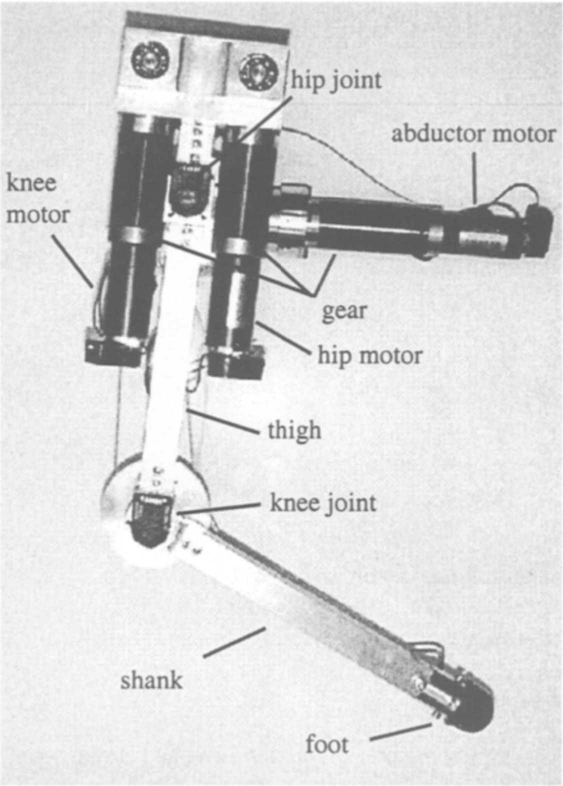


Figure 1. Side view of the M₃L leg. The total length and weight is 67 cm and 6.8 kg respectively. The length of the thigh is 24 cm and the shank 30 cm.

The range of motion of the leg can be seen in figure 2. The abduction joint has a range of -30 degrees

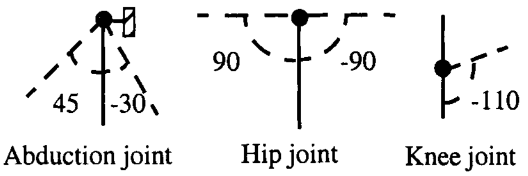


Figure 2. Range of motion of the leg

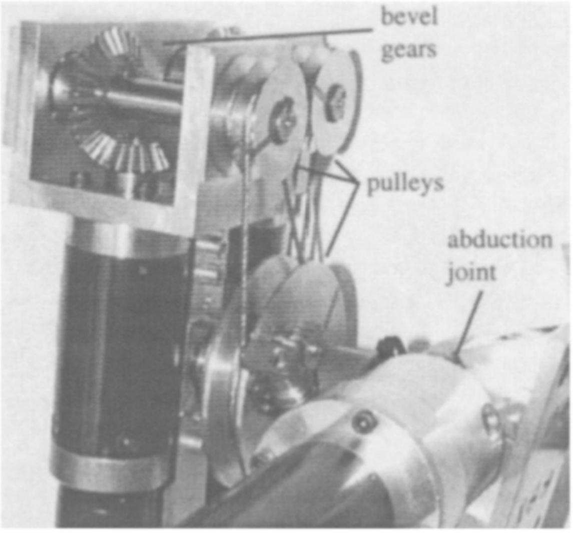


Figure 3. Inside view of the hip. The bevel gears change the axis of rotation from the vertical to horizontal and rotation is translated to the joints with the cables and pulleys. The abduction joint rotates around a horizontal axis, perpendicular to the axis of rotation of the pulleys.

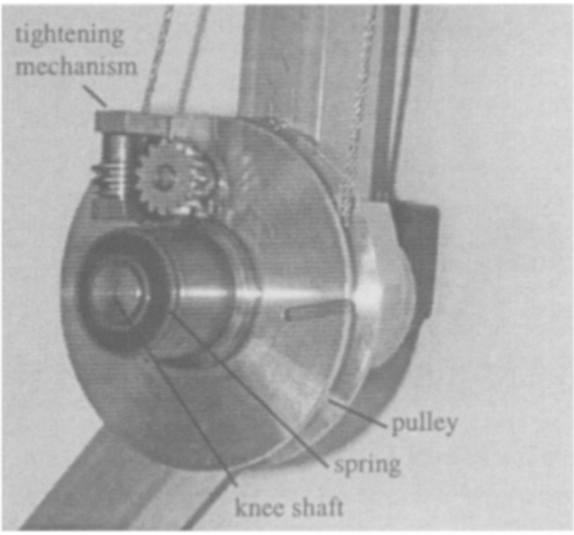


Figure 4. Close-up of the knee joint showing the implementation of the series elasticity. The spring is made of two concentric steel cylinders with rubber between them. The spring is pressed on the knee shaft and the pulley is then pressed on the spring. The tightening mechanism for the cables can also be seen.

underneath the robot and 45 degrees outside measured from vertical, the hip joint has -90 to 90 degrees and the knee joint has 0 to -110 degrees measured relative to the thigh where 0 is full extension of the knee. The motion of all the joints is limited by mechanical stops.

The weight is reduced by using aluminium where possible, with the thigh and shank having hollow rectangular profiles. To reduce the moment of inertia of the leg, the motors and respective gears for the hip and knee joints are placed close to the body and balanced around the hip joint, aligned with the thigh, see figure 1. Three 90 Watt Maxon DC motors were chosen along with planetary gears. Bevel gears and a system of steel cables and pulleys are then used to transfer rotation to the joints, see figure 3. To minimize the risk of the cables getting caught in obstacles, the pulleys and the cables are placed on the inside of the leg.

The reduction ratio for the gears in the hip and knee joints is 180:1 where 36:1 is provided by a two stage planetary gear, 2.5:1 by a bevel gear and 2:1 by pulleys. The abduction motor is mounted horizontally and connected, via a 216:1 three stage planetary gear, directly to the abduction joint. The maximum continuous speeds and torques for these choices of motors and reduction ratios are 4.8 rad/s and 19.3 Nm respectively for the hip and knee joints and 4.0 rad/s and 23.1 Nm respectively for the abduction joint.

The leg has a rubber pad as a foot with an integrated one-axial force sensor. The total weight of the leg is 6.8 kg where the weight of the moving parts are 4.8 kg. The total length is 67 cm, the length of thigh is 24 cm and the shank is 30 cm, where the proportions between thigh and shank were based on the proportions of a hindlimb of a cat.

3.2. Series elasticity

To mimic the elastic properties of muscles, rubber torque springs are implemented in series with the actuators in the hip and knee joints, see figure 4. The use of springs smooth out the impact forces as has been mentioned before but additionally they reduce the peak torques that have to be taken by the gears, thereby increasing their life-time. Silentblocks are used as torque springs which are made of two concentric cylinders with rubber injected between them. The spring constant is approximately 116 Nm/rad for the rubber spring and the total spring constant with the cables is approximately 72 Nm/rad and is fairly linear for the range of torques that will be applied. A spring was not implemented in the abduction joint as the motion of the joint was judged to be not periodic enough to be able to reutilize the potential energy stored in the spring. It would though have given the joint more shock tolerance if the spring had been implemented.

In order to measure the deflection of the springs in the knee and hip joints, two encoders are placed on each joint, one at motor and one at the joint itself. This makes it possible to measure the position of the motor and joint and also to get an estimate of the torque in the joint.

4. EXPERIMENTAL EVALUATION

This chapter will present an experiment that was done to evaluate the legs characteristics and performance against impacts. The leg is attached to a wagon that is able to move along a vertical rail. A treadmill allows the leg to walk or run although it is not used in the following experiments. Current controlled motordrivers were chosen as they enable the controller to give torque references to the motors. The leg is connected to a dSpace lab card with a in-built DSP, sitting in a PC computer, for running the controller and do the signal processing.

The leg was dropped from a height of 60 cm measured from ground to abduction joint. The leg was slightly bent and the foot was approximately 10 cm above the ground. A controller proposed by Hogan [7] is used to implement springlike behavior between body and foot in cartesian coordinates. The control law is

$$T = J^T(q_m)[K_c(X_0 - L(q_m)) - D_c J(q_m)\dot{q}_m] \qquad (2)$$

where T is the controlled actuator torque, q_m is the motor angles, X_0 is the desired end-point position in cartesian coordinates, K_c and D_c are diagonal stiffness and damping matrixes for the controller, L is the kinematic relation between end-point position in cartesian coordinates and joint positions and J is the Jacobian. By letting the controller work only on motor positions, the controller and the joint springs work in series. The reference foot position to the controller was 50 cm in the vertical direction and 0 cm in the horizontal measured from the hip joint during the whole experiment. The controller stiffness is chosen much higher than the spring stiffness so most of the compliance during the impact comes from the springs. If this is compared to equation (1), the torque around each joint can be controlled but the stiffness is constant and dependent on the rubber springs' properties. It should be stressed that this controller is only for the purpose of studying the response of the joint springs.

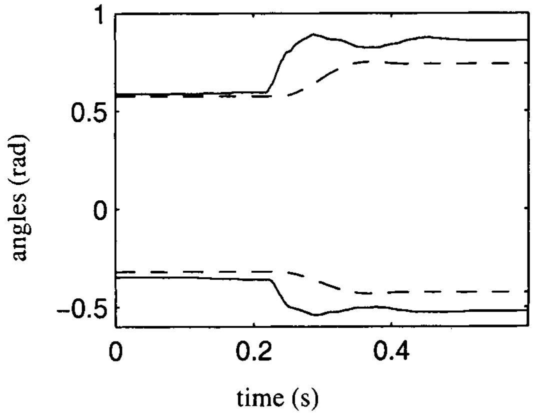


Figure 5. Joint angles (solid) and motor angles (dashed). Upper is for the knee and lower for the hip

Measured joint trajectories can be seen in figure 5 and show a smooth motion of the joints and the motor position. The motor approaches its new equilibrium position with only a small overshoot. The commanded torques and spring torques can be seen in figure 6 where the spring torques are simply calculated by

$$T_S = k(q_m - q) \qquad (3)$$

where T_s is the torque in the spring, k is the spring constant, q_m is the motor position and q is the joint position.

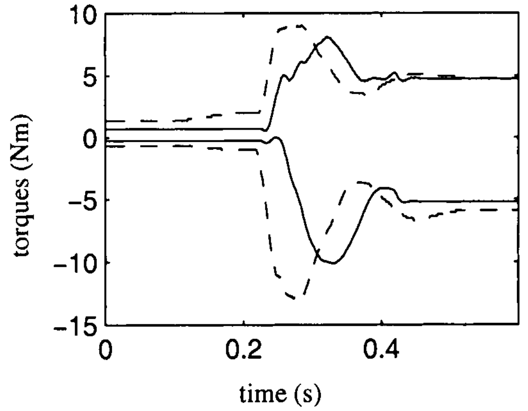


Figure 6. Controlled (solid) and spring torques (dashed). Upper is for the hip and lower for the knee

sition. As can be seen the build up of torque in the springs is gradual and has none of the characteristics of an impact. This gives the controller some extra time to respond. The peak value of the torque in the springs is larger than what the controller needs to give which reduces the energy consumption. It is also worth noticing the potential energy stored in the springs in figure 6.

One observation that was made, especially when the leg was dropped from higher heights, was that the foot tended to jump or jitter at touch down. This could be avoided by increasing the compliance of the controller but that would result in an increased motion of the joints and thereby more energy spent. A better approach would be to redesign the foot making it more able to damp and absorb the impact than the current rubber pad.

5. CONCLUSIONS AND FUTURE WORK

The design of the M3L has been described where it was found that biology can offer inspiration to the design. The leg uses DC-motors in series with torsional rubber springs to mimic muscular properties of legged animals. It has a light, efficient and compact design and uses revolute joints which makes it possible to take advantage of the pendular properties of the thigh and shank. The characteristics of the springs were evaluated experimentally and it was found that the springs reduce the necessary response time and the maximum torque that the controller has to give, and give a smooth motion of the joints.

Future work will include implementation of a control concept where the use of the potential energy storage in the torsion springs and the passive properties of the leg will be investigated.

ACKNOWLEDGMENTS

This work was supported by the Foundation for Strategic Research through its Centre of Autonomous Systems at the Royal Institute of Technology, Stockholm, Sweden.

REFERENCES

[1] R. McN. Alexander, Three uses of springs in legged locomotion, Int. J. of Rob. Res., vol. 9, no. 2,1990

[2] K. Benjelloun, Design and construction of the M_3L robot leg, Technical report, TRITA-MMK 1997:21, Dept. of Machine Design, Royal Inst. of Technology, Stockholm, Sweden, 1997

[3] R.J. Full, Integration of individual leg dynamics with whole body movement in arthropod locomotion, Biological neural networks in invertebrate neuroethology and robotics,Academic press, 1993

[4] M. Garcia, A. Chatterjee, A. Ruina, M. Coleman, Complex Behavior of the Simplest Walking Model, Dept. of Theoretical and Applied Mechanics, Cornell University, 1996

[5] S. Hirose, A study of design and control of a quadruped walking machine, Int. J. Robot. Res. 3:113-133, 1984

[6] N. Hogan, Adaptive control of mechanical impedance by coactivation of antagonist muscles, IEEE Trans. on Autom. Control, vol. ac-29, no. 8, 1984

[7] N. Hogan, Impedance control: An approach to manipulation: Part I-III, J. of dyn. syst., measurements and control, 107:1-24, 1985

[8] H. Kazerooni, Compliant motion control of robot manipulators, Int. J. Control, vol. 49, no. 3, 1989

[9] G. Mennitto and M. Buehler, CARL: A compliant articulated leg for dynamic locomotion, Robotics and autonomous systems 18:337-344, 1996

[10] G.A. Pratt, M.M. Williamson, P. Dillworth, J. Pratt, K. Ulland and A. Wright, Stiffness isn't everything, Preprints of the fourth symposium on experimental robotics (ISER '95), 1995

[11] J. Pratt, P. Dillworth and G.A. Pratt, Virtual model control of a bipedal walking robot, Proc. 1997 IEEE Int. Conf. Robot. and Autom., 1997

[12] R.D. Quinn and K.S. Espenschied, Control of a hexapod robot using biologically inspired neural network, Biological neural networks in invertebrate neuroethology and robotics,Academic press, 1993

[13] M.H. Raibert, Legged robots that balance, Cambridge, Massachusetts: MIT Press, 1986

[14] S.M. Song and K.J. Waldron, Machines that walk: The adaptive suspension vehicle, Cambridge, Massachusetts: MIT Press, 1989

[15] T. Wadden and Ö. Ekeberg, A neuro-mechanical model of legged locomotion: single leg control, to be published in Biological Cybernetics

[16] J. Yamaguchi and A. Takanishi, Development of a biped walking robot having antagonistic driven joints using nonlinear spring mechanism, Proc. 1997 IEEE Int. Conf. Robot. and Autom., 1997

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Adaptive Neural Control of Hexapod Leg Trajectories

M. J. Randall, A. G. Pipe and A.F.T. Winfield

Intelligent Autonomous Systems Engineering Laboratory, Faculty of Engineering,
University of the West of England, Coldharbour Lane, Bristol, BS16 1QY, United Kingdom

In this paper we describe our initial results from the first phase of an investigation into on-line adaptive neural control of a hexapod walking robot's leg trajectories. In the intended application area very accurate trajectories, and foot placement, are required - even in the presence of disturbances. Our first experiments build upon existing experience in neural control of robot manipulators. The first experiments are concerned with the neural controller's performance in the face of intermittent end effector force disturbances.

1. INTRODUCTION

This paper describes the lowest layer implementation of our control architecture for a hexapod robot [1-3].

1.1. Trajectory Generation Strategy

The third layer of our architecture as described in the companion paper [1] provides the leg trajectories for the hexapod. Defining a leg trajectory is an ill-posed problem since there are an infinite number of possible trajectories that the foot could follow. In this research, we considered some results of applied psychology studies of human arm trajectories. There are two reasons for choosing such a starting point. First, it is intended that the artificial hexapod reduce "wear and tear" on its mechanical components by employing some appropriate cost-minimising function observed in biological systems. Second, it is intended to generate trajectories that appear "natural", as a human observer of the robot would perceive it. In addition, the studies provide many insights into some human cognitive level organisational principles behind trajectory determination. In fact, the "aesthetic" features of natural movements have been shown to correspond to just such organisational principles or rules.

In our work, leg trajectory planning is accomplished in kinematic terms using the extrinsic foot co-ordinates rather than joint co-ordinates. For a full justification see [2]. In order to do this, it is necessary to use inverse kinematics [3]. These planned movements are executed by effectors with highly non-linear dynamics (whether they be muscles or robotic actuators), and in biological systems many neural networks and neural pathways are responsible for executing such movements. This makes understanding the computational level of such movements highly non-trivial. Thus, while the kinematic planning of trajectories will be expressed in endpoint co-ordinates, at the lowest level of the control architecture, adaptive neural controllers are implemented that compensate for systematic dynamic uncertainties at the level of the individual actuators.

One interesting observation concerning human arm trajectories, which serves a useful purpose later, will be mentioned here. The Isochrony Principle [4] states that the duration of movements for large and small segments of the motion is roughly equal. The Isochrony Principle has been shown to apply to both free and constrained movements. In the context of this work, it is convenient to define a related principle, which we shall call the "Strong Isochrony Principle". The Strong Isochrony Principle is defined as follows: "For movements segmented by via points, the durations of movement along the paths represented by each segment and terminated by predefined via points are equal throughout the whole movement". The effect of the Strong Isochrony Principle is to constrain the movement spatially and temporally simultaneously.

1.2. Neural Control Research in Our Lab

The recent experiments reported here have been carried out on one front leg of a hexapod robot. However the neural control approach used is heavily based in the previous research conducted in the Intelligent Autonomous Systems engineering Laboratory (IAS) at the University of the West of England (UWE) on joint level neural control of robot manipulators. It is therefore necessary to review salient findings from this work in order to set the current experiments in context.

For some years, IAS researchers have been working on intelligent neural control and its applications. The example application of high precision trajectory control in revolute multi-jointed robot manipulators was chosen early on because of its appropriate level of complexity as a cross-coupled non-linear MIMO (Multi-Input Multi-Output) system. However, this work can be set in a wider context of establishing a theory, and thereby crucial design constraints, for the application of on-line learning neuro-control to a range of plant displaying complex non-linear dynamics. In particular, one of the principal interests has been in tackling those problems wherein the plant dynamics possess significant components which are difficult (sometimes even impossible) to model by more traditional control theoretic approaches. This might be because mathematics at a sufficient level of complexity becomes intractable or because the required data to build an analytical model is difficult to obtain. These are quite common problems when tackling high precision control problems with a complex plant. However it should be stated that without these constraints it is difficult to justify straying from a traditional control theoretic approach (although there is a body of relevant published research work which does not seem to take this into account).

After early experimentation the team spent some time concentrating on establishing the necessary stability theory from which to construct design rules for the application in mind. After this, implementations were demonstrated for improving trajectory tracking accuracy on a simulated Puma 560 robot manipulator and on a real educational Mentor robot [5-6]. More recently we have spent time extending the experimental work to cover a real Yaskawa Motaman robot [7], extracted from daily use as a manufacturing welding robot in a local company.

In the latter case it was possible to report a six- to ten-fold reduction in joint level trajectory tracking errors on each of the principal three axes of movement compared with using the Yaskawa's existing joint controllers on their own. This was achieved by adding our neuro-controller architectures to those of the existing robot. It was possible to utilise the existing sensory feedback signals directly and to complement the original joint control signals either by analogue summation close to the drive motors or digitally at an earlier stage.

2. MINIMUM JERK TRAJECTORIES

In order to generate a unique protraction trajectory for the hexapod leg, we have chosen to minimise its jerk, after [8], where "jerk" is defined as the time derivative of acceleration. Strictly speaking, the mean-squared jerk is minimised over the total duration of a movement. Our minimum jerk objective function is an extension to a three-dimensional description of the swing trajectory and is as follows:

$$C_F = \frac{1}{2}\int_{t_0}^{t_f}\left(\left(\frac{d^3 p_x}{dt^3}\right)^2 + \left(\frac{d^3 p_y}{dt^3}\right)^2 + \left(\frac{d^3 p_z}{dt^3}\right)^2\right) dt \qquad (1)$$

where: $p_x(t)$, $p_y(t)$ and $p_z(t)$ define the foot trajectory in three dimensional Cartesian co-ordinates relative to the coxa-thoracic joint as in (1), and the subscript "F" stands for foot.

It is necessary to constrain the swing trajectory through a "via-point". The mathematical details of this become complex and so those interested are referred to [2] where the exact form of the swing trajectory as predicted by dynamic optimisation theory is presented. The support phase does not require a via-point and we use the method of [9].

Several assumptions have been made in the trajectory generator. The first is that the biological data presented by [10-11] for stick insects is suitable for the hexapod. This is a valid assumption since the robot design is based on the stick insect with similar dimensions (albeit on a different scale). The second assumption that has been made concerns the nature of the gait. For the experiments presented here, each phase of the trajectory is separated into 100 time steps. This is compatible with a tripod gait that has no delay between phase transitions (i.e. from stance to swing or from swing to stance). In order to turn the demand trajectory into a series of planned actuator signals, the range of Pulse-Position-Modulation (PPM) signals to the leg servos was calibrated and a suitable transformation applied.

3. STABLE ON-LINE LEARNING USING ADAPTIVE NEURAL CONTROL

3.1. The Yaskawa Robot Application

For illustrative purposes let us consider the more recent Yaskawa robot experimental situation in a little more detail, after all there are some significant similarities (as well as differences) between a revolute jointed industrial robot manipulator and each leg of the hexapod robot of prime interest here. One could consider the three major axes (base, shoulder, elbow) of an industrial manipulator as a coupled system of links and revolute joints, driven by electric motors. The Yaskawa is fitted with a shaft encoder and tachometer on each of these joints so as to provide position and velocity feedback which may be used by a joint level controller.

The general form of a rigid manipulator can be written as

$$\mathbf{H}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}} + \bar{\mathbf{g}}(\mathbf{q}) + \mathbf{F}(\mathbf{q},\dot{\mathbf{q}}) = \tau(\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}}) \qquad (2)$$

where: $\mathbf{q}$ is the n x 1 vector of joint displacements, τ is the $n\times1$ vector of applied joint torques, $\mathbf{H}(\mathbf{q})$ is the $n\times n$ symmetric positive definite manipulator inertia matrix, $\mathbf{C}(\mathbf{q},\dot{\mathbf{q}})\dot{\mathbf{q}}$ is the $n\times1$ vector of centripetal and Coriolis torques, $\bar{\mathbf{g}}(\mathbf{q})$ is the $n\times1$ vector of gravitational torques, and $\mathbf{F}(\mathbf{q},\dot{\mathbf{q}})$ are the unstructured uncertainties of the dynamics including friction and other disturbances. Although the static performance of the PID controllers, commonly used in such situations, is good, the dynamic performance leaves much to be desired. By exploiting the Neural Network universal approximation feature we can assume that the left hand side of the above equation can be approximated by a neural network, i.e.

$$\mathbf{H}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \bar{\mathbf{g}} + \mathbf{F} = \mathbf{G}(\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}})^{\mathrm{T}}\mathbf{W} + \xi(\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}}) \qquad (3)$$

where: $\mathbf{G}$ is the neural network nonlinear mapping matrix, $\mathbf{W}$ is the neural network weight vector which is initially unknown, and ξ is the neural network approximation error which should be as small as possible by careful design.

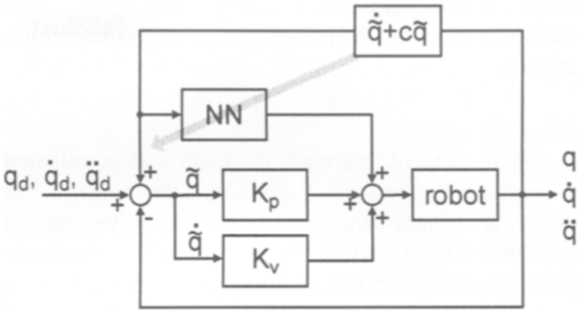


Figure 1 Yaskawa neural control architecture

In the experimental approach adopted for the Yaskawa robot a CMAC (Cerebellum Model Articulation Controller) [12] neural network is effectively "strapped around" the existing controller, rather than replacing it; though the theoretical work does encompass complete replacement. Three neural controllers are used, one for each of the three major axes. They are implemented via a PC-resident Texas Instruments DSP board, based around the 40MHz 320C40 processor, which executes the code for all three neural networks. Each neural network utilises the same demand and feedback information available to the existing controller, and the output signals from the existing and neural controllers are combined by simple analogue voltage addition just before the power stages of the motor drive circuit. Each neural network therefore simply adds its control effort to that of the existing controller in order to help reduce errors. Over repeated trials performance improves; though the bulk of this is obtained in the first five trials of a given trajectory.

The arrangement for one joint is depicted in figure 1. Each composite joint controller receives new position, velocity and acceleration demands every 2ms. The existing two term linear controller operates on only the position and velocity error signals to derive control actions, whilst the neural network takes position, velocity and acceleration demands as inputs and is trained on-line using a combination of velocity and position errors.

In the last year these experimental results have been extended by testing the composite controller in many more parts of the manipulator's operational envelope, enabling an empirical analysis of its learning and generalisation performance across many trajectories [7]. The results of this work have been encouraging, including one or two recommendations for enhancement of the controller implementation.

3.2. Discussion

By the end of this work the team had established a sound theoretical basis for the design of *convergent* and *stable* on-line adaptive neural controllers for a certain class of complex non-linear problem demonstrated on the three major axes of a real industrial welding robot. The principal aim is to design systems capable of improving performance by modelling crucial components of a given plant directly using the data normally available as part of the control process; e.g. control effort, desired and actual set-points, their derivatives and so on.

Off-line techniques involve two main stages. First large amounts of data are typically obtained from operating the plant with some simpler controller, or from simulations. This gathered information is then used as training data for a neural network or fuzzy logic system. After training, a fixed transfer function has been learned "by example" which can then be used in improving performance. However there are some limitations which can be overcome by an *on-line* approach wherein the processes of data gathering and training occur simultaneously with control of the real plant. Firstly the off-line approach suffers from a "Catch-22" situation when only plant simulations are used to derive training data; if the simulation is sufficiently complex to accurately model the complex dynamics then this information could probably have been used to build a more traditional controller. Secondly, even if real plant data are used then the learned transfer function is limited to that information which has been acquired during the data gathering phase, thereby disallowing adaptation to untested plant dynamics or unexpected time variations.

However an *on-line* approach, as developed by researchers at UWE, is not without difficulties. If the controller characteristics are changing due to on-line adaptation as the real plant is exercised, then it becomes imperative to provide guarantees of learning algorithm convergence and stability of the whole system. Yichuang Jin, an IAS group member, noticed at the beginning of his research that there was little published work in this area and so the main thrust of his PhD thesis quickly became that of establishing a set of design rules which would satisfy these constraints [5].

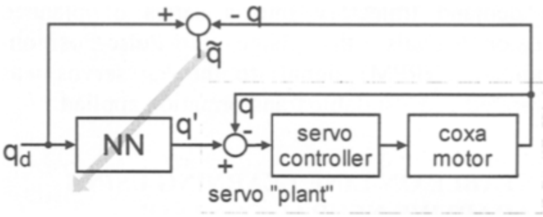


Figure 2 The leg joint neural control scheme

3.3. Related Work

Over the years much research effort has been put into the design of neural network applications for manipulator control. Albus [12] used the Cerebellum Model Articulation Controller (CMAC) to control manipulators as early as 1975. Miller et al [13] and Kraft et al [14] extended Albus' results and developed neural network learning algorithms in 1987 and 1992 respectively. Kawato et al [15] added MLP networks to original manipulator PD control systems as feedforward compensators in 1988. Iiguni et al [16] combined manipulator linear optimal control techniques with Multi-Layer Perceptron (MLP) neural networks which were used to compensate the nonlinear uncertainty in 1991. Many aspects of this groundbreaking work were crucial in the field for a number of reasons. More recently, others working in the emergent field of Intelligent Control have concentrated on linking novel design proposals and experimental results with learning convergence and stability guarantees. For example, Lewis and Jagannathan [17] have done this for a range of applications, and control structures that include *off-line* trained neuro-controllers and *on-line* learning cases.

4. APPLYING ADAPTIVE NEURAL CONTROL TO A HEXAPOD LEG JOINT

In figure 2, we identify a simplified neural control structure (relative to the Yaskawa experiments described above) which implements control of the coxa joint of one of the hexapod's front legs; later this will be extended to control of the other two joints of each leg. A Radial Basis Function (RBF) neural network is used as controller [18]. It is trained off-line initially and is then operated with an on-line learning algorithm. This arrangement represents the current state of ongoing work; first experimental results are given in the following section. Possible improvements and further work are identified in the final concluding section.

In the Yaskawa robot experiments, the neural network contributed directly to control current supplied to each joint's motor. In this case each motor is packaged as a servo-system, with on-board linear position control. A desired position is identified by Pulse-Position-Modulation (PPM). To generate a trajectory, desired positions are produced in sequence for each of the three servo-motors of a leg by the on-board computer (described in [1]). In these experiments the RBF network is placed between this computer output and the servo-motor input, as detailed in figure 2.

The RBF network is trained initially with an identity function so that, without on-line learning, it contributes no extra control effort. An error signal is derived from the position potentiometer already contained in the servo as part of the existing position control loop. This feedback signal is read into the on-board computer via an A/D converter and then used to derive a position error signal. This signal is used in turn to train the neural network on-line between one control action and the next, thus generating a modified trajectory demand signal which should reduce trajectory tracking errors in the face of external disturbances. In the Yaskawa experiments velocity feedback was used in addition to position. This will be required here also when the experimental platform is extended. For the initial results presented here, however, only one desired trajectory is considered and so velocity variations can be ignored.

5. EXPERIMENTAL RESULTS

All figures show position against time in the integer units of the trajectory-planning algorithm, and in time steps of approximately 20 ms. There are three plots on each graph. The "bold white" plot is the desired position for the coxa joint, the "grey" plot is the servomotor position feedback signal, and finally the "bold black" plot is the RBF neural network output.

Figure 3 shows the initial off-load case on the first run through the trajectory. It can be seen that all three signals are quite similar under these conditions, meaning that the servo-controller is performing quite well under these circumstances and

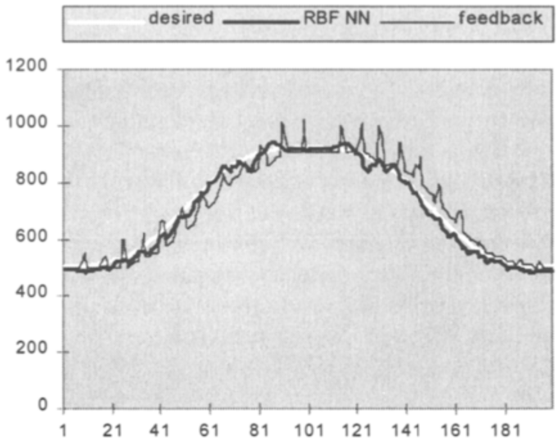


Figure 3 The initial desired, RBF output and feedback trajectories

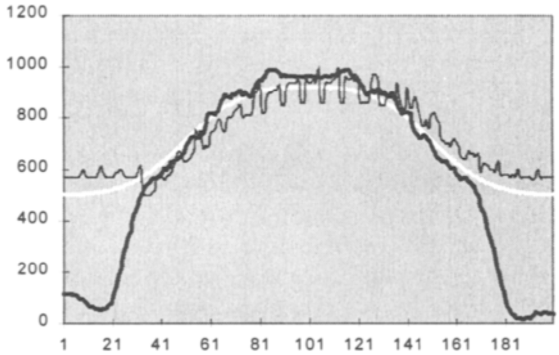


Figure 4 The desired, RBF output and feedback trajectories after adapting to the spring.

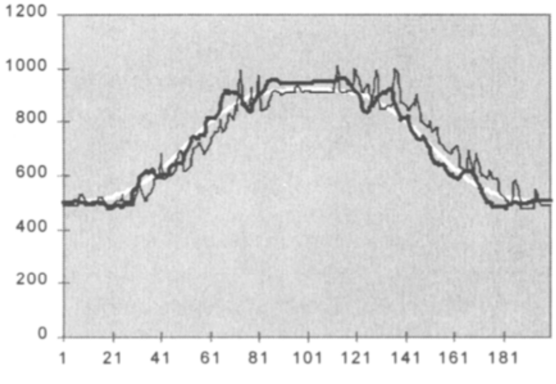


Figure 5 The desired, RBF output and feedback after adapting to the removal of the spring.

the RBF network is unnecessary.

Figure 4 shows the situation at the fourth cycle of the trajectory, with a spring attached to the leg end-effector, which applies a variable tangential load to the system. We can clearly see that the RBF neural network has learnt to compensate for tracking inaccuracies over the preceding three cycles, producing a modified trajectory demand which is substantially different from the original so as to give the required position feedback signal.

After this the spring was removed. This produced a significant initial trajectory overshoot. However by the seventh cycle the situation was nearly recovered to the original situation, as shown in figure 5.

6. DISCUSSION AND CONCLUSIONS

These results show promise for further work. There are a number of problems to be overcome. Perhaps most significant is the noise generated by the potentiometer based feedback signal from the servo-system, this can be seen clearly in all the diagrams above. This requires either complete replacement, or perhaps some filtering.

We would like to discard the internal servo-position control loop and replace this with direct current control from the RBF neural network output. It may be possible at this stage to replace the position (and generate new velocity) feedback sources in the space thus created inside the servo-mechanism.

These experiments require extension to other leg joints, and other legs, of the hexapod. This will involve careful computer processing load, and power requirements, analyses.

REFERENCES

1. M. J. Randall, A.G. Pipe, A.F.T. Winfield, 'An Intelligent Mechatronic Architecture for Hexapod Control' (same volume).
2. M.J. Randall, A.G. Pipe, Proc. Intl Workshop Adv Robotics and Intelligent Machines 1997.
3. M.J. Randall, A.F.T. Winfield, A.G. Pipe, IEE Digest 96/167 (1996).
4. P. Viviani, C. Terzuolo, Neuroscience, 7(2), 431-437, (1982).
5. Y. Jin, PhD Thesis 'Intelligent Neural Control and Its Applications in Robotics', University of the West of England, UK. (1994)
6. Y Jin, A G Pipe, A Winfield, 1997, "Chapter 5: Stable Manipulator Trajectory Control Using Neural Networks", *Neural Systems for Robotics*, Eds. Omid Omidvar & Patrick van der Smagt, Academic Press, ISBN 0-12-526280-9.
7. R. P. Cherian, "On-Line Learning Neuro-Control of an Industrial Manipulator", MSc Thesis, Engineering Faculty, UWE, Bristol, UK, 1997.
8. T. Flash, N. Hogan, J Neuroscience, 5(7) 1688-1703, (1985)
9. E.A. Devjanin, V.S. Gurfinkel, E.V. Gurfinkel, V.A. Kartashev, A.V. Lensky, A. Yu. Shneider, L.G. Shtilman, Mechanism and Machine Theory, 18(4), 257-260, (1983).
10. H. Cruse, Biol Cybernetics, 24, 25-33, (1976).
11. H. Cruse, Ch. Bartling, J. Insect Physiol., 41(9), 761-771, (1995).
12. J. S. Albus, J of Dynamic Systems, Measurement, & Control, 97, pp.220-227, 1975.
13. W.T. Miller III, F.H. Glanz, L.G. Kraft III, International Journal of Robotics Research, Vol.6, No.2 pp.84-98, 1987.
14. L.G. Kraft, W.T. Miller, and D. Dietz, Handbook of Intelligent Control, Neural, Fuzzy, and Adaptive Approaches, eds: D.A.White and D.A.Sofge, New York: Multiscience Press Inc., pp.215-232, 1992.
15. M. Kawato, Y. Uno, M. Isobe, and Suzuki, R., IEEE Control System Magazine, 8(2), pp.8-15, 1988.
16. Y. Iiguni, H. Sakai, and H. Tokumaru, IEEE Trans. on Neural Networks, 2, pp.410-417, 1991.
17. F. W. Lewis, S. Jagannathan, 'Neural Network Control of Robot Manipulators and Non-linear Systems', Taylor and Francis (to appear 1998).
18. R.M. Sanner, and J.J.E. Slotine, IEEE Trans. on Neural Networks, 3, pp.837-863, 1992.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Linear viscoelastic actuator-based control system of a bipedal walking robot

V. Berbyuk[a], B. Peterson[b] and N. Nishchenko[a]

[a]Pidstryhach Institute for Applied Problems of Mechanics and Mathematics,
Ukrainian National Academy of Sciences,
3-B, Naukova St., 290601, Lviv, Ukraine

[b]Department of Mechanics, Chalmers University of Technology,
S-412 96, Gothenburg, Sweden

The optimisation approach for designing of the rotational spring-damper actuators providing the programmed goal-directed motion of a bipedal walking robot is proposed. The problem is formulated as an approximation procedure for the controlling torques acting at the joints of the robot during its optimal motion. Analysis of the obtained numerical results has shown that the anthropomorphic energy-optimal goal-directed motion of the bipedal walking robot could be generated by the rotational spring-damper actuators with one switching of each of their parameters during the double step of the robot.

1.INTRODUCTION

The complexity of the structure and the dynamical instability of the gait of the bipedal walking robots (BWR) make it very difficult to understand the main features of their control systems needed to provide the goal-directed stable motion. A number of mathematical models of the bipedal locomotion systems with different degrees of complexity have been proposed during last two decades [1-11]. These models were used for different purposes, e.g. to design the anthropomorphic mechatronic system (exoskeleton) [1], to study human normal gait [3,8] and pathological gait [8,9], to develop the control systems for the BWR [1-7], etc.

Nevertheless, presently there is not enough information about how to create an effective control system of the BWR, especially to provide its extremal, e.g. time-optimal or energy-optimal motion. What type of controlling forces of the actuators are required for these extremal controlled motion of the robots?

The central idea of this paper is that an effective control system of a mobile robot can be designed based on a combination of passive and active controlling stimuli. The passive actuators have to provide the programmed goal-directed motion of the robot. The active controlling stimuli should be used for some other aims, e.g. to solve the stabilisation problem of motion of the robot.

The optimisation approach for designing of the rotational spring-damper actuators providing the programmed goal-directed motion of a bipedal walking robot is proposed. The problem is formulated as the approximation procedure of the controlling torques acting at the joints of the robot during its optimal motion. This motion and respective torques are determined by solution of the optimal control problem for the dynamical system modelling the robot. In the optimal control problem the objective function has been chosen as the integral over a double step of the sum of the absolute values of the mechanical power of all controlling torques acting at the joints of the robot [5-7]. The restrictions on the phase coordinates and the controlling forces have been generated based on experimental data of human locomotion [8]. In order to solve the nonlinear optimal control problem an approach based on the inverse dynamics method and the Rosenbrock's algorithm has been used [10].

Analysis of the obtained numerical results has shown that the anthropomorphic energy-optimal goal-directed motion of the bipedal walking robot could be generated by the rotational spring-damper actuators with one switching for each of their parameters during the double step of the robot.

2.MATHEMATICAL MODEL

The BWR under consideration consists of an inertial torso with weight and two identical legs (Fig.1). Each leg consists of three elements. The two elements with mass and rotatory inertia model the thigh and the shank, while a third massless and inertia-free element models the foot.

Let the system OXYZ be a fixed rectangular Cartesian coordinate system. It is assumed that the robot moves in the OXY plane along the OX axis, over a horizontal surface.

In addition to the weight of the torso, thighs and shanks, the external forces acting on the BWR include the interaction forces between the feet and the surface, which are replaced by the force $\mathbf{R}_i(t)$ applied at the point of the ankle joint Ai, and the moment $p_i(t)$ that acts in the same joint (i=1,2). The force $\mathbf{R}_i(t)$ is the principal vector of the plane system of reaction forces of the support; $p_i(t)$ is the principal moment of the reaction forces of the support, referred to the point of the ankle joint. It is also assumed that the control torques $q_i(t)$ and $u_i(t)$ (i=1,2) are acting at the hip H and knee Ki joints, respectively.

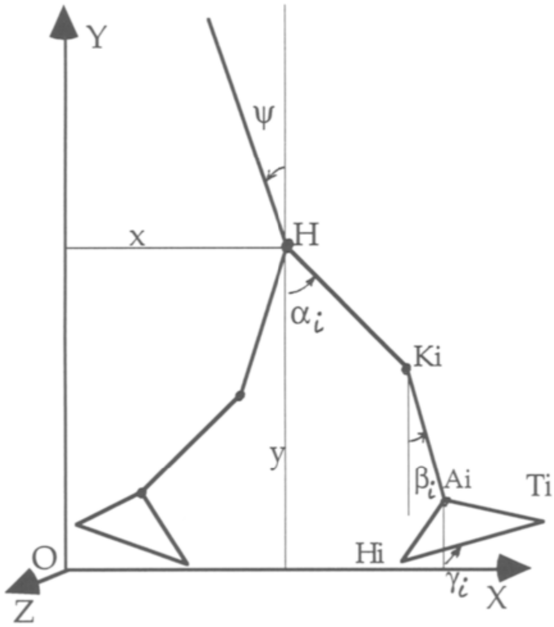


Fig.1. Model of the Bipedal Walking Robot

The controlled motion of the BWR can be described by the Lagrange's equations of the second kind and the kineto-static balance conditions for the feet under the action of the ankle moment and the reaction of the support [7,9,10]:

$$\begin{aligned}
&f_1 - K_r(\ddot{\psi}\cos\psi - \dot{\psi}^2\sin\psi) = R_{1x} + R_{2x}\\
&f_2 - K_r(\ddot{\psi}\sin\psi + \dot{\psi}^2\cos\psi) = R_{1y} + R_{2y}\\
&f_{3i} = q_i - u_i + a(R_{ix}\cos\alpha_i + R_{iy}\sin\alpha_i) \qquad (1)\\
&f_{4i} = u_i - p_i + b(R_{ix}\cos\beta_i + R_{iy}\sin\beta_i)\\
&f_5 = -q_1 - q_2\\
&p_i + (y_i - y_{Ri})R_{ix} - (x_i - x_{Ri})R_{iy} = 0,
\end{aligned}$$

where

$$\begin{aligned}
&f_1 = M\ddot{x} + \sum_{i=1}^{2}[K_a(\dot{\alpha}_i\cos\alpha_i)^{\cdot} + K_b(\dot{\beta}_i\cos\beta_i)^{\cdot}]\\
&f_2 = M(\ddot{y} + g) + \sum_{i=1}^{2}[K_a(\dot{\alpha}_i\sin\alpha_i)^{\cdot} +\\
&K_b(\dot{\beta}_i\sin\beta_i)^{\cdot}]\\
&f_{3i} = J_1\ddot{\alpha}_i + K_a(\ddot{x}\cos\alpha_i + \ddot{y}\sin\alpha_i) +\\
&aK_b[\ddot{\beta}_i\cos(\alpha_i - \beta_i) + \dot{\beta}_i^2\sin(\alpha_i - \beta_i)] +\\
&gK_a\sin\alpha_i, \quad (i = 1,2)\\
&f_{4i} = J_b\ddot{\beta}_i + K_b(\ddot{x}\cos\beta_i + \ddot{y}\sin\beta_i) +\\
&aK_b[\ddot{\alpha}_i\cos(\alpha_i - \beta_i) - \dot{\alpha}_i^2\sin(\alpha_i - \beta_i)] +\\
&gK_b\sin\beta_i, \quad (i = 1,2)\\
&f_5 = J\ddot{\psi} - K_r(\ddot{x}\cos\psi + \ddot{y}\sin\psi) - gK_r\sin\psi\\
&M = m + 2m_a + 2m_b, \qquad K_a = m_a r_a + a m_b\\
&K_b = m_b r_b, \qquad K_r = rm, \qquad J_1 = J_a + a^2 m_b
\end{aligned}$$

In the above expressions: $x, y, \psi, \alpha_i, \beta_i, \gamma_i$ are the Cartesian coordinates of the suspension point H of the legs and the angles that specify the position of the elements of the BWR (Fig.1); m, r, J are the mass, the distance from the suspension point of the legs to the centre of mass and the moment of inertia of the torso relative to the Z axis at point H, respectively; m_a, r_a, a, J_a are the mass, the distance from H to the centre of mass, the length and the moment of inertia of the thigh relative to the Z axis at H, respectively; m_b, r_b, b, J_b are the mass, the distance from Ki to the centre of mass, the length and the moment of inertia of the shank relative to the Z axis at the point Ki, respectively; R_{ix}, R_{iy} are the horizontal and vertical components of the force $\mathbf{R}_i(t)$; $(x_i, y_i), (x_{Ri}, y_{Ri})$ are the Cartesian coordinates of the ankle joint and of the point of application of the vector $\mathbf{R}_i(t)$ of the i-th leg, respectively; g is the acceleration due to gravity, and the dot " $\dot{}$ " is a derivative with respect to time.

3. STATEMENT OF THE PROBLEM

The object under consideration is a nonlinear multidimensional controlled dynamical system. The problem of designing the passive actuators is formulated as the approximation procedure of the controlling torques acting at the joints of the robot during its energy-optimal motion.

Let us describe some kinematic characteristics of the typical gait of the BWR before stating of the respective optimal control problem.

We shall assume that there are four phases of the leg action during a double step of the robot ($t \in [0,T]$): the *first* double ($t \in [0,t_1]$) and single ($t \in [t_1,t_2]$) support phases, and the *second* double ($t \in [t_2,t_3]$) and single ($t \in [t_3,T]$) support phases [6-8]. This rhythm of the double step of the robot leads to the following kinematic constraints:

$$
\begin{aligned}
&x_{H1}(t) = x_{H1}^0, \quad y_{H1}(t) = 0, \quad t \in [0,t_1]\\
&x_{T2}(t) = x_{T2}^0, \quad y_{T2}(t) = 0, \quad t \in [0,t_1] \qquad (2)\\
&x_{T1}(t) = x_{H1}^0 + l, \quad y_{T1}(t) = 0, \quad t \in [t_1,t_3]\\
&x_{H2}(t) = x_{H1}^0 + L, \quad y_{H2}(t) = 0, \quad t \in [t_2,t_3]\\
&x_{T2}(t) = x_{T2}^0 + 2L, \quad y_{T2}(t) = 0, \quad t \in [t_3,T]\\
&x_{H1}(T) = x_{H1}^0 + 2L, \quad y_{H1}(T) = 0,
\end{aligned}
$$

where $x_{H1}^0, x_{T2}^0, t_1, t_2, t_3, T$ are given parameters; l is the length of the foot, L is the length of the single step; $(x_{Hi}, y_{Hi}), (x_{Ti}, y_{Ti})$ are the Cartesian coordinates of the heel and toe of the i-th leg, respectively. These coordinates are determined by the expressions (Fig.1):

$$
\begin{aligned}
&x_{Hi}(t) = x_i - l_1\cos(\gamma_i - \varphi)\\
&y_{Hi}(t) = y_i - l_1\sin(\gamma_i - \varphi)\\
&x_{Ti}(t) = x_i + l_2\sin(\gamma_i - \varphi)\\
&y_{Ti}(t) = y_i - l_2\cos(\gamma_i - \varphi) \qquad (3)\\
&x_i(t) = x + a\sin\alpha_i + b\sin\beta_i\\
&y_i(t) = y - a\cos\alpha_i - b\cos\beta_i\\
&\varphi = \arctan(l_1 / l_2)\\
&l_1 = A_iH_i, l_2 = A_iT_i, \angle H_iA_iH_i = \pi / 2.
\end{aligned}
$$

Consider the following optimal control problem. *Problem A.* It is required to determine the state vector $\mathbf{z(t)}$ of the robot and the vector of the controlling stimuli $\mathbf{u(t)}$ which both satisfy the equations of motion (1), the kinematic constraints (2),(3), the boundary conditions

$$
\begin{aligned}
&\mathbf{z}_x(T) = \mathbf{z}_x(0), \quad x(T) = x(0) + 2L \qquad (4)\\
&\mathbf{z}_x = \left\{\dot{x}, y, \dot{y}, \psi, \dot{\psi}, \alpha_i, \dot{\alpha}_i, \beta_i, \dot{\beta}_i, \gamma_i, \dot{\gamma}_i\right\},
\end{aligned}
$$

the restrictions on the phase coordinates and controlling stimuli

$$
\begin{aligned}
&y(t) \geq h, \quad \alpha_i(t) \geq \beta_i(t), \quad t \in [0,T]\\
&y_{H1}(t) > 0, \quad y_{T1}(t) > 0, \quad t \in [t_3,T]\\
&y_{H2}(t) > 0, \quad y_{T2}(t) > 0, \quad t \in [t_1,t_2]\\
&R_{iy}(t) \geq 0, \quad t \in [0,T], \quad i = 1,2\\
&x_{H1}^0 \leq x_{R1}(t) \leq x_{H1}^0 + l, \quad t \in [0,t_2]\\
&x_{R1}(t) = x_{H1}^0 + l, \quad t \in [t_2,t_3]\\
&x_{R2}(t) = x_{T2}, \quad t \in [0,t_1] \qquad (5)\\
&x_{H1}^0 + L \leq x_{R2}(t) \leq x_{H1}^0 + L + l, \quad t \in [t_2,T]\\
&\gamma_2(t) - \beta_2(t) - \pi / 2 = \Theta_{a2}(t), t \in [0,T]\\
&\alpha_2(t) - \beta_2(t) = \Theta_{k2}(t), \quad t \in [0,T]\\
&\alpha_i(t) - \psi(t) = \Theta_{hi}(t), t \in [0,T], \quad i = 1,2\\
&R_{2x}(t) = R_{2x}^e(t), \quad t \in [0,t_1] \bigcup [t_2,t_3]\\
&R_{2y}(t) = R_{2y}^e(t), \quad t \in [0,t_1] \bigcup [t_2,t_3],
\end{aligned}
$$

and which minimise the functional [6,7,9,10]:

$$
\begin{aligned}
E = \frac{1}{2L}\int_0^T \sum_{i=1}^{2} [|q_i(\dot{\psi} - \dot{\alpha}_i)| + |u_i(\dot{\alpha}_i - \dot{\beta}_i)| + \qquad (6)\\
|p_i(\dot{\beta}_i - \dot{\gamma}_i)|]dt.
\end{aligned}
$$

Here the functions $\Theta_{a2}, \Theta_{k2}, \Theta_{hi}, R_{2x}^e, R_{2y}^e$, ($i$=1,2) are given from the experimental data of human locomotion [8]. The boundary conditions (4) and constraints (2),(3),(5) are imposed on the phase state and controlling stimuli of the BWR with the aim to approach its controlling motion to human gait as close as possible.

The problem of designing the rotational spring-damper actuators of the BWR is formulated as follows.

Problem B. Let $\mu(t)$ be the control torque acting at some joint (μ-joint) of the leg during the energy-optimal motion of the BWR. It is required to determine the vector ξ_μ, which minimise the functional

$$\Phi_\mu = \int_0^T [\mu - w(t,\xi_\mu)]^2 dt, \qquad (7)$$

where

$$w = \begin{cases} d_{\mu 1} + c_{\mu 1}\varphi_\mu(t) + k_{\mu 1}\dot{\varphi}_\mu(t), t \in [0, t_\mu] \\ d_{\mu 2} + c_{\mu 2}\varphi_\mu(t) + k_{\mu 2}\dot{\varphi}_\mu(t), t \in [t_\mu, T] \end{cases} \qquad (8)$$

$$\xi_\mu = \{d_{\mu j}, c_{\mu j}, k_{\mu j}, t_\mu, j = 1,2\}$$

$$\mu = \{q_i^*(t), u_i^*(t), p_i^*(t), i = 1,2\}.$$

In formulae (8): $c_{\mu j}, k_{\mu j}$ are the torsional spring and damping coefficients of the actuator acting at the μ-joint, respectively; $\varphi_\mu(t), \dot{\varphi}_\mu(t)$ are the angle between the links connected by the μ-joint and the corresponding angular velocity during the energy-optimal motion of the BWR; $t_\mu, 0 \le t_\mu \le T$ is the time of switching of the actuator's parameters; $d_{\mu j}$ is determined by the free angle of the spring and torsional spring coefficient, (j=1,2).

The function (8) with the vector ξ_μ which minimise the functional (7) will determine the linear viscoelastic actuator of the μ-joint of the BWR.

4. RESULTS AND DISCUSSION

The proposed approach for the design of the rotational spring-damper actuators-based control system of the BWR requires the solution of *Problems A* and *B*. The algorithm of the numerical solution of *Problem A* is developed [10] using the special Fourier and spline approximation of the independent variable functions and inverse dynamics method. Below all numerical results are presented for the following values of the linear and mass-inertia parameters of the BWR: m, r, J are equal to 46.7 kg, 0.39 m and 7.1 kgm^2, respectively; m_a, r_a, a, J_a are equal to 7.08 kg, 0.16 m, 0.41 m and 0.082 kgm^2, respectively; m_b, r_b, b, J_b are equal to 5.04 kg, 0.203 m, 0.5 m and 0.053 kgm^2, respectively; l_1=0.05 m, l_2=0.2 m. These values of the parameters of the BWR correspond to the respective parameters of human body with total mass M=73.2 kg and height of 1.76 m [11].

Some results of the solution of *Problem A* for the gait with natural cadence (T=1.1196 s, L=0.755 m) [8] are shown in Fig.2-10 (dashed curves).

Figure 2 shows the way in which the knee angle $\Theta_{k1}(t) = \alpha_1(t) - \beta_1(t)$ of the leg of the robot changes in time over a double step for the obtained energetically optimal law of motion.

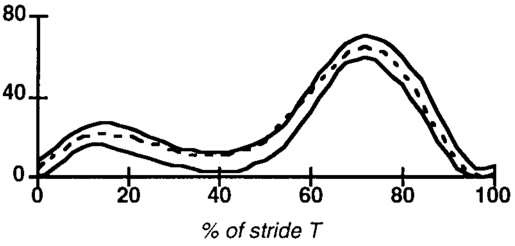


Fig.2 Knee angle $\Theta_{k1}(t)$, in degrees

The way in which the specific horizontal component $R_{1x}(t)/M$ of the support reaction varies (Fig.3, dashed curve) indicates that in each single step the support leg successively executes two functions: deceleration of the robot (time interval in which $R_{1x}(t)$/M<0) and separation (time interval in which $R_{1x}(t)/M$>0). The maximum value of the $R_{1x}(t)$ amounts to 20% of the entire weight of the robot. The vertical component of the support reaction ($R_{1y}(t)/M$, Fig.4, dashed curve) exceeds the weight of the robot by not more then 7%.

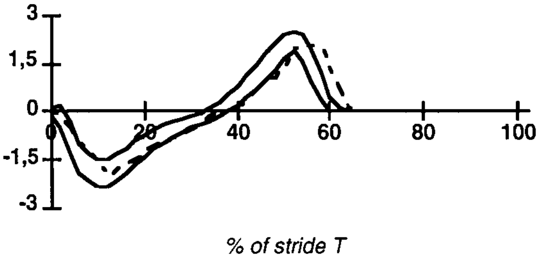


Fig.3 Horizontal force $R_{1x}(t)/M$, in N/kg

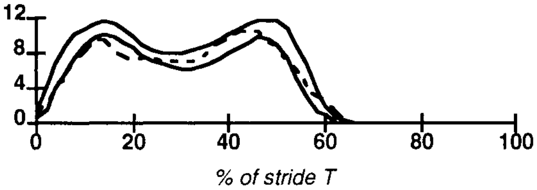


Fig.4 Vertical force $R_{1y}(t)/M$, in N/kg

For comparison purposes in Fig.2-4 the domains of the values of the respective characteristics obtained by experiments for human normal gait are shown (the domains are bounded by the solid curves) [8]. The analysis of these data and all above mentioned indicate that the kinematic and dynamic characteristics of the obtained energy-optimal law of motion of a BWR are within reasonable proximity to the corresponding characteristics of human gait.

Figures 5-10 show the specific control torques $q_i^*(t)/M, u_i^*(t)/M, p_i^*(t)/M, i=1,2$ (dashed curves) acting at the joints of the legs during the energy-optimal law of motion of the robot.

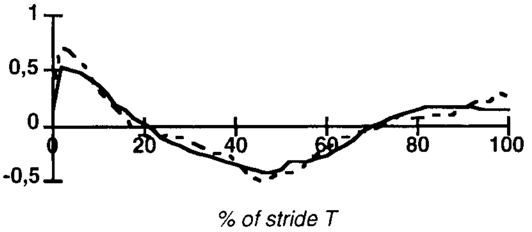


Fig.5 Hip torque $q_1^*(t)/M$, in Nm/kg

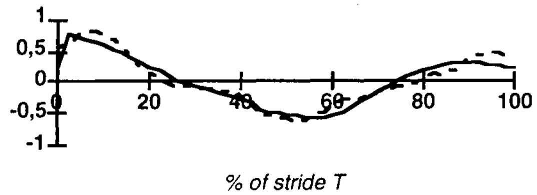


Fig.6 Hip torque $q_2^*(t)/M$, in Nm/kg

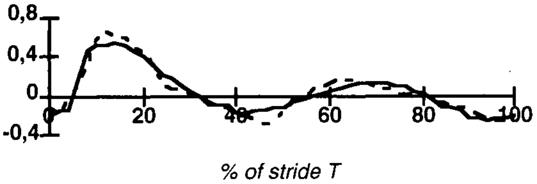


Fig.7 Knee torque $u_1^*(t)/M$, in Nm/kg

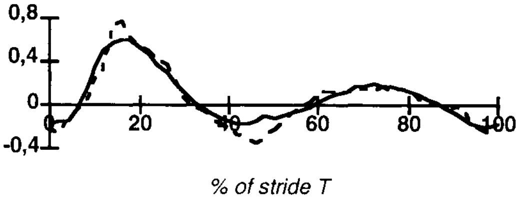


Fig.8 Knee torque $u_2^*(t)/M$, in Nm/kg

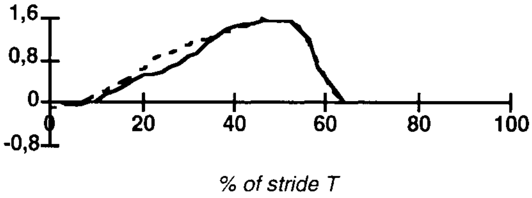


Fig.9 Ankle torque $p_1^*(t)/M$, in Nm/kg

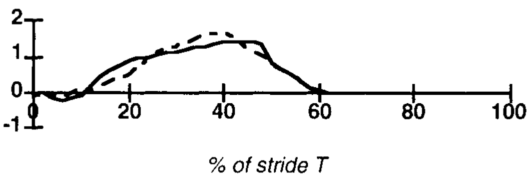


Fig. 10 Ankle torque $p_2^*(t)/M$, in Nm/kg

The solution of *Problem B*, that is the vectors $\xi_\mu = \{d_{\mu j}, c_{\mu j}, k_{\mu j}, t_\mu, j=1,2\}$, for these optimal control torques are determined by data of the Table.

The graphic dependencies of the viscoelastic approximation of the controlled torques for the energy-optimal motion of BWR are depicted in Figures 5-10 (solid curves). The analysis of these graphic dependencies shows that the control stimuli required for the energy-optimal motion of the BWR can be constructed with sufficient accuracy by the torsional spring-damper actuators. It is also can be noted that these control stimuli quantitatively and qualitatively are within reasonable proximity to the corresponding characteristics of human locomotion [8]. This is additional evidence of fruitfulness of the proposed approach for modelling of bipedal walking.

The analysis of the data of Table shows that within the framework of our assumption the spring coefficients of the designed energy-optimal actuators are one-two order of magnitude larger then the damping coefficients. The stiffness parameters of the actuators of ankles joints of the BWR are the greatest and are approximately 2-3 times greater then those of the actuators of hip joints.

Table

Torsional spring (Nm/rad) and damping (Nms/rad) coefficients of the actuators and time (%*T*) of their switching

	$q_1^*(t)/M$	$q_2^*(t)/M$	$u_1^*(t)/M$	$u_2^*(t)/M$	$p_1^*(t)/M$	$p_2^*(t)/M$
$c_{\mu 1}$	1.821	2.249	3.361	2.991	5.361	5.224
$k_{\mu 1}$	0.180	0.103	0.048	0.090	0.082	0.003
$d_{\mu 1}$	-0.013	0.059	-0.745	-0.553	0.289	0.248
$c_{\mu 2}$	0.886	1.482	0.319	0.336	3.411	1.835
$k_{\mu 2}$	0.003	0.049	0.011	0.002	0.002	0.066
$d_{\mu 2}$	0	0	0	0	0	0
t_μ	48	46	44	46	36	36

5. CONCLUSION

In this paper the analysis of the goal-directed motion of the BWR is based on the solution of energy-optimal control problem for a plane multibody system with nine degrees of freedom. Taking into account the obtained results the following conclusions can be drawn.

The proposed optimisation approach for designing of the rotational spring-damper actuators based control system of a BWR is fruitful.

The anthropomorphic energy-optimal motion of the BWR could be generated by the rotational spring-damper actuators with one switching for each of their parameters during the double step of the robot.

6. ACKNOWLEDGEMENTS

This work was partially supported by The Wenner-Gren Center Foundation and by The Swedish Institute, Stockholm, Sweden.

We are also greatful for the kind hospitality of professors Anders Boström and Peter Olsson at the Department of Mechanics, Chalmers University of Technology, Göteborg, Sweden.

REFERENCES

1. M. Vukobratovich, *Walking Robots and Anthropomorphic Mechanism,* Mir Press, Moscow, 1976.
2. V.B. Larin, *Control of walking apparatuses,* Naukova Dumka, Kiev, 1980.
3. H. Hatze, Neuromusculoskeletal control systems modeling - a critical survey of recent developments, *IEEE Transactions on Automatic Control,* **AC-25** 5 (1980) 375-385.
4. A.M. Formal'sky, *Displacement of anthropomorphic mechanisms,* Nauka, Moskow, 1982.
5. V.V. Beletskii, V.E. Berbyuk and V.A. Samsonov, Parametric optimisation of motion of a bipedal walking robot, *J. Mechanics of Solids,* **17** (1982) 24-35.
6. V.V. Beletskii, *Two-Legged Walking: Model Problems of Dynamics and Control* , Nauka, Moscow, 1984.
7. V.E. Berbyuk, *Dynamics and Optimisation of Robototechnical Systems*, Naukova Dumka, Kiev, 1989.
8. D. Winter, *The Biomechanics and Motor Control of Human Gait,* University of Waterloo Press, Canada, 1991.
9. V. Berbyuk, Dynamic and optimal control problems for biotechnical systems "Man-Prosthesis". *In D.H. van Campen (editor), IUTAM Symposium on Interaction Between Dynamics and Control in Advanced Mechanical Systems,* Kluwer Academic Publishers, The Netherlands, (1997) 35-43.
10. V. Berbyuk, G. Krasyuk and N. Nishchenko, Mathematical modelling of the human gait dynamics in sagittal plane, *J. Matematuchni Metody i Fizyko-Mechanichni Polya,* Ukrainian National Academy of Sciences, **40** 4 (1997) 65-76.
11. A. Cappozzo, T. Leo and A. Pedotti, A general computing method for the analysis of human locomotion, *J. Biomechanics*, **8** (1975) 307-320.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Simulation and Emulation of Sensor Systems for Intelligent Vehicles

Sandor Ujvari[*], Patric Eriksson[¤], Philip Moore[✤] and Junsheng Pu[✤]

[*]Centre for Intelligent Automation, Department of Engineering Science
Högskolan Skövde
S-541 28 Skövde, Sweden

[¤]Research Division, Prosolvia Systems AB
S-462 21, Vänersborg, Sweden

[✤]Mechatronics Research Group, Faculty of Computing Sciences and Engineering
De Montfort University
Leicester, LE1 9BH, UK

Abstract: Simulation of sensor systems for mobile robots are described in this paper. By simulation of smart sensor systems, the performance of semi-autonomous vehicles / mobile robots can be enhanced. Smart sensor systems used in the field of mobile robotics can utilise adaptive algorithms. e. g. artificial neural nets, fuzzy logic or hybrid variants of these systems. The development, training and evaluation of adaptive algorithms for sensor systems can be done within a virtual environment in which graphical models are built to simulate an intelligent vehicle, its sensors, and its environment. The virtual sensors are validated by comparing the characteristics of the virtual sensors with those of the real devices.

1.0 INTRODUCTION

The movement of materials within a production process has a very important function in manufacturing. One way to meet the needs for materials handling in industrial applications such as automotive manufacture and assembly is the use of assembly carriers and guided mobile robots. These systems need to be highly flexible to cope with changing production schedules and transportation routings. The most common sensors used for navigation in these systems are wire-guidance techniques, guidance beacons or visible marker lines on the floor.

By introducing 'smart sensors' that have a higher degree of sophistication, higher level functionality can be achieved, either as a complement, or replacement for the basic navigation methods. An area of Virtual Manufacturing / Digital Plant Technology is Computer Aided Robotics (sometimes referred to as Virtual Robotics) where graphical workstations can be used to design/visualise, model and simulate the operation of robotic and automation installations (e.g. based on 3-D CAD systems), leading to the generation of control programmes for the real machines. Facilities can be derived from such 'virtual design' environments, whereby virtual sensor components can be incorporated. Such a virtual environment can be used to support the evaluation and pre-emptive learning/training of semi-autonomous vehicles / mobile robots which use sensors and reactive behavioural based control [1]. Control code can be developed for the sensor systems through simulation of sensors, the vehicle, and the environment within which the vehicle operates. The possibility to design and train, through pre-emptive learning, control algorithms for smart sensors in a virtual environment can facilitate proof-of-concept and dramatically reduce the need for on site training [2].

The support which can be provided by a virtual environment during the design or re-configuration of a mobile vehicle includes; i) development of new control algorithms for sensors; ii) integration of sensors with the control architecture; and iii) evaluation of vehicle performance. A sensor-configuration which already exists can be tested in a new environment, whereby the new environment is modelled and the modelled vehicle can be exposed to this unknown environment without jeopardising the vehicle or the environment.

Since the sensors are often combined so the mobile robot can solve a task, the problem of adapting / designing the sensor systems for a specific task can be difficult and time-consuming. This can be the case if artificial neural nets or fuzzy logic algorithms are used, where the training of ANNs and the designing of FL is difficult to accomplish on a real vehicle due to safety and time consideration. The computer aided robotics tool Cimstation Robotics has been used to develop an environment for the simulation of mobile vehicles and sensor systems. Experiments are presented of simulations of assembly carriers with wire-guidance sensor and laser range scanner.

1.1 CONTROL ARCHITECTURES FOR MOBILE ROBOTS

Traditionally control architectures for mobile robots are decomposed into functions, so called functional decomposition. Brooks proposed a different approach, where the architecture is decomposed into task achieving modules, also known as horizontal decomposition [9].

There are several levels of links perception-action, and task can be low level as such as wall following, or higher planning level, e. g. navigate through an unknown area. Many approaches and techniques have been used to achieve the connection between perception and action for tasks, as neural networks [10] and fuzzy logic [11]. When developing and testing techniques and algorithms in virtual environments, a library of sensor models can be used, with a number of sensors and sensor systems modelled at a high level of accuracy, and a well defined interface between the sensors and the control system. The sensors can then be configured according to characteristics, with the overall task of the vehicle in mind, and control algorithms can either be taken from the library, developed or both. Control algorithms for lower level tasks, as wall following, can be developed using the sensor information from several sensors. In the wall following example a FL algorithm can be used, taking into account both i) the relatively accurate readings from the laser range scanner (a picture presenting the position and function of sensor systems), however only in one plane just above floor level, and ii) the less accurate readings of an ultrasonic sensor system that covers the 3D space in front of the vehicle. The algorithms are then transferred to the real vehicle. This solution should shorten the time to develop applications from a given vehicle, and to customise the vehicle to a new environment or task. Material handling system application can in this way be built for plants only existing as drawings, new functions developed with either existing or new configurations of sensors.

2.0 MOBILE ROBOT SENSORS

The perception of mobile robots is critical to fulfil navigation tasks. Everett states that ‘Fundamental in this regard are the required sensors with which to acquire high-resolution data describing the robot’s physical surroundings in a timely yet practical fashion , and in keeping with the limited onboard energy and computational resources of a mobile vehicle.’ [7]

Sensor system design is complex and not easily conveyed from a theoretical perspective only. When considering sensor simulation, the validation of models is therefore highly important. In this context, a tool that provides beforehand answers to how a certain sensor system, combination of systems, or a vehicles performance in a new environment, can be of great use, provided that the model limitations are not exceeded.

3.0 SENSOR MODELLING

Some of the work made in modelling sensors using 3D-graphical computer tools are the SAVIC simulator developed by Chen et al [3,4], where several sensor types are modelled, laser scanner, ultra sonic sensors etc. The characteristics of a proximity sensor is approximated by using a three dimensional line segment. The sensor is then simulated by checking collisions of the line segments with other objects in the environment.

The sensor models in this work are based on a method whereby the sensing, or detection shape of the sensor is represented as a geometrical shape (e.g. a CAD-model attached to the sensor-body) [2]. By performing collision detection between the detection shape and other geometrical shapes in the environment (obstacles etc.), it can be decided what is detected, figure 1. For range measuring sensors, the model can be extended with so called trace-lines, see figure 6. These are used to provide the distance to the detected objects along each of the lines. The simulated sensors are independent devices that have their own control programmes determining their behaviour. Sensor characteristics such as detection range, output format and up-date frequency are very important aspects of the model, and necessary for a true correlation between the real device and the model.

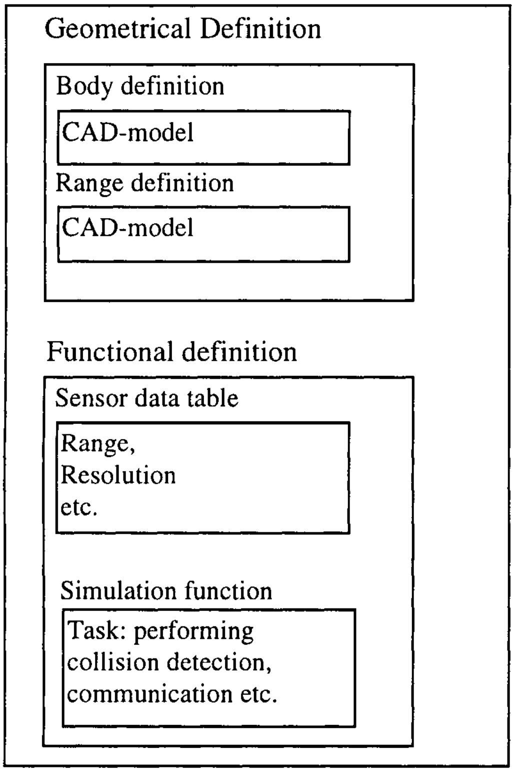


Figure 1. A generic sensor model proposed by Eriksson [1]. The physical description of the sensor are CAD-models, which include the detection shape.

By changing the functional definitions in the generic sensor model, it is possible to adjust the measurement rate and the detection range to correspond with the parameters of a real sensor. The simulation functions are specific for simulating the interaction of the sensor and the environment.

3.1 Wire guidance sensor

A commonly used navigation method for materials handling applications is wire guidance. Such system typically detects a magnetic field generated around the wire-path by alternating current in the kHz range. The sensor, or antenna, provides an analogue output signal proportional to the distance between the wire and the antenna. There are other variants of guide-path following using optical stripes or magnetic tapes [7]. The apparent disadvantage of these methods is the lack of flexibility, the advantage though is the robustness of the system. The idea of moving away from fixed path guidance is growing in popularity trough out the AGV industry, where combinations of fixed guide-path following and free-ranging AGVs can be found.

Due to similarities, the wire-guidance model described can with little change emulate optical stripe or magnetic tapes. The wire guidance sensor model is made using the generic sensor model, and resembles the antenna in many aspects, i) the same output characteristics, and timing, ii) Same detection distance iii) same internal operation using two coils (see fig 2). The wire path is made of line and curved segments of some shape. The function of the sensor is simulated by using two coils from which the shortest vector to the active wire segment is calculated. An output signal is emulated by the wire sensor model.

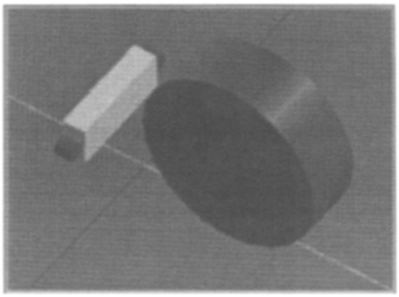


Figure 2. The frontwheel with a wiresensor with two coils attached. The coils are sensing the electromagnetic field around the wire. The difference in signal between the left and the right coil controls the steering of the wheel via a PD-regulator.

A= distance from right coil
B= distance from left coil
X= Output signal from sensor

$$\frac{A - B}{A + B} = X$$

Figure 3. Output function from the wire sensor. The effect of the signal magnitude on the output signal is removed Below a threshold level of A and B, the output is set to zero. Consequently, a negative value of X gives that the wire is to the right of the sensor.

The sensor has a control program which is executed at discrete time steps. The length of a time step is determined by the update rate of the sensor (see figure 5). The program checks for collisions between the detection shape and the wire path segments. If any collisions occur, the shortest vector from each coil to the collided path segment is calculated. Finally the output signal is formatted and stored as a global variable.

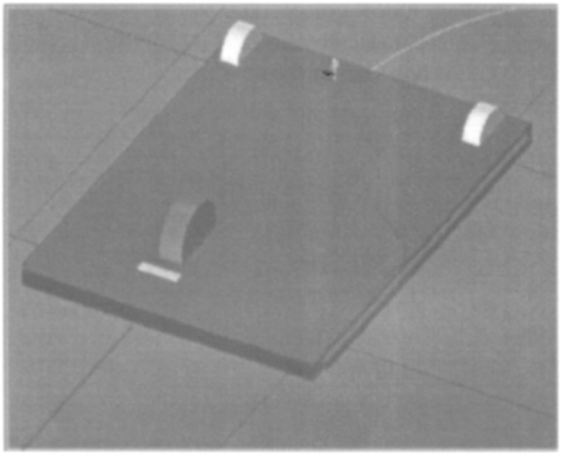


Figure 4, The base platform without the chassi. The front wheel is driving and steering, where the steer signal comes from a wiresensor, in front of the front wheel. (The vehicle can be seen in figure 7).

When the sensor control process is activated, it works independently of other vehicle control processes, and if a wire is within range, it will provide the distance. The model operates regardless of the direction of the wire, if several wire segments are present, the closest one is chosen.

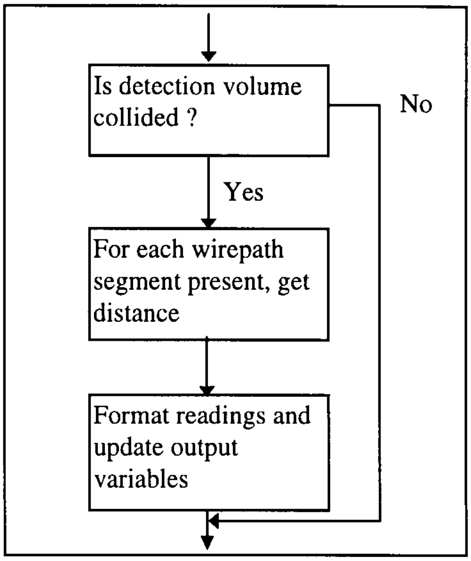

Is detection volume collided ?
No
Yes
For each wirepath segment present, get distance
Format readings and update output variables

Figure 5. Output function from the wire sensor, which is performed each update interval.

3.2 Laser range scanner model

The second sensor to be modelled is a laser range scanner using the principle of time-of-flight, made by Sick AG[5]. The scanning is performed in one plane, and the field of view is 180 degrees. Range is under ideal conditions up to 50 metres, with a resolution +/- 50 mm and 0.5 degrees. The laser range scanner can be used for safety purposes, e. g. as a front looking safety device, replacing bumpers [6].

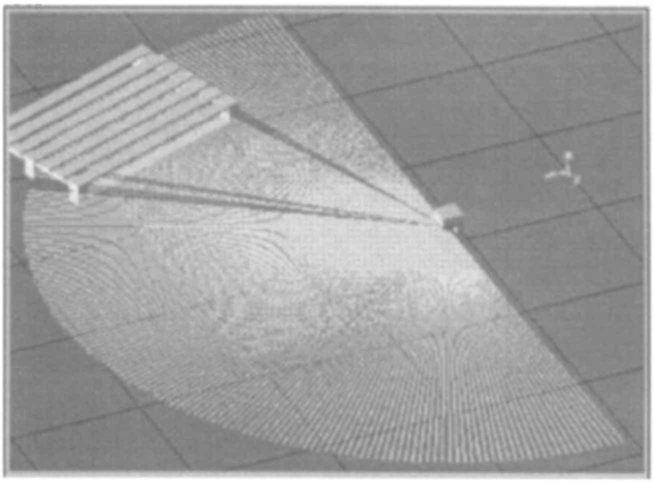


Figure 6. Laser range scanner with 181 trace-lines detecting a wooden pallet. The colliding lines change colours.

The model is made, according to the generic sensor model, with one geometrical sensor body representation, and 181 detection lines, so called tracelines. There are two additional shapes in the detection range representation, the warning field and the protection field. When the warning field collides, a signal is sent, usually resulting in decreased speed, while collisions with the protection field is equivalent with an emergency break. At every update of the sensor output (80 milliseconds), the closest colliding object is found for each traceline.

5.0 EXPERIMENTS AND RESULTS

The experimental vehicle used in this study is equipped with a wire-guidance sensor, a laser range-scanner for obstacle detection, and shaft encoders on the drive wheel for dead-reckoning. The vehicle is described closer in [9] It can perform wire following and free range motion through dead reckoning.

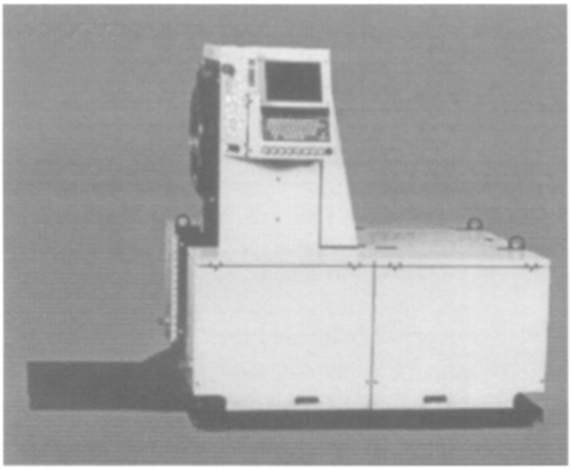
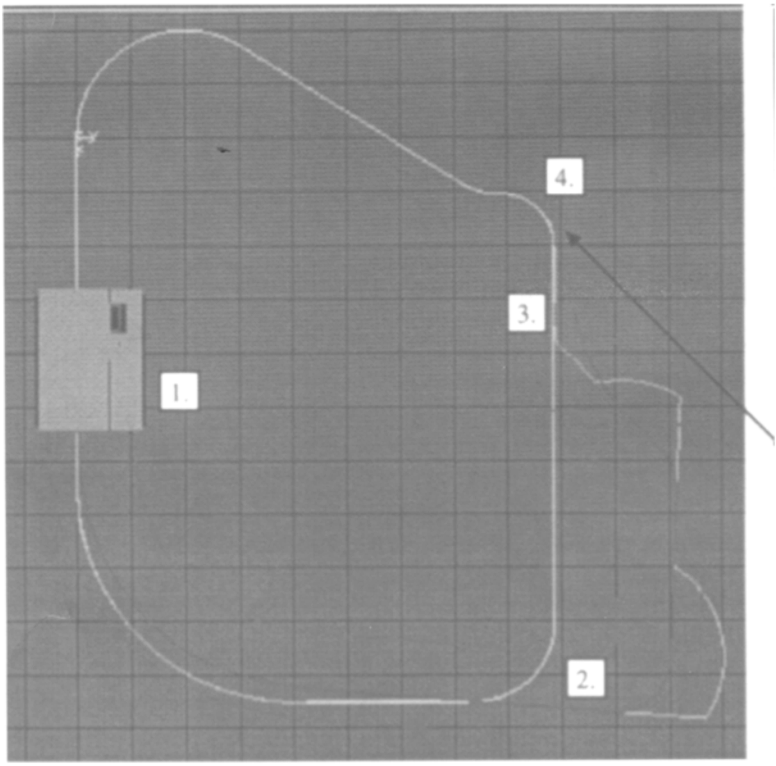


Figure 7. The test vehicle, an assembly carrier (Volvo Automation).

The laser range scanner was used as a safety device The speed of the vehicle is set as percentage of maximum speed.



Test of wire following function
The testruns have been performed using the vehicle, the wiresensor an the laser range scanner models.
The model vehicle executes the same motion commands as the real vehicle. During the travel, the wire sensor updates the distance to the wire every 30 millisecond. If the speed when turning corners is too high, the vehicle will drive off the wire, which happened to the model and the real vehicle at the same turn.

1. Vehicle in start position, and starts to follow wire.
2. When a certain distance is run, the vehicle leaves the wire and runs in free ranging mode.
3. When the wire is detected, wire following continues
4. The vehicle moves too far from the wire, drops it and execution is halted.

Figure 8. Orthogonal perspective from above. The carrier is in starting position, and the lines show the trajectory, the white line is the wire path. The tiles are 0.5*0.5 meters. The carrier has started from a position to the left in the picture, travelled along the wire, then executed an off-wire motion command, making two turns to the left. Eventually the vehicle found the path again, but lost the wire due to too high velocity at the sharp turn.

The wire guide path is used for testing AGVs after being assembled. The tests showed a close resemblance between performance of sensor models and real devices.
A virtual environment has been used to accurately simulate an assembly carrier in an event based simulation, using two sensors operating as independent devices, independent of other processes in the simulation environment.

The environment can be used for:

i) the validation of sensor models comparing emulated and real device characteristics,
ii) the development and training of sensor control algorithms, pre-emptive learning
iii) testing of performance against defined measures.

It has been shown that sensors of highly different types can be modelled using the generic sensor model, i. e. wire guidance sensor and a laser range scanner, that these can be emulated to a high level of accuracy operating concurrently. A control architecture can be tested utilising the information from the sensor models.

Future work will be to look into the development of an ultra-sonic sonar system, sensor fusion, the development of algorithms, artificial neural networks and fuzzy logic approaches, in the virtual environment, and transfer procedures for the transfer of off-line developed code onto the real vehicle.

ACKNOWLEDGEMENTS

The research programme described in this paper is funded by the INCO-Copernicus programme IC15CT960726 and The Swedish National Board for Industrial and Technical Development, NUTEK. The project involves major contributions from Volvo Automation (Sweden), Kaunas University of Technology (Lithuania), TCC and MSTU Bauman (Russia).

REFERENCES

1. Moore P R, and Eriksson, P., A Role for 'Sensor Simulation' and 'Pre-emptive Learning' in Computer Aided Robotics, Proceedings of 26th International Symposium on Industrial Robots (ISIR'95), Singapore, 4-6 October, 1995.
2. Eriksson, P and Moore, P. R., An Environment for Off-Line Programming and Simulation of Sensors for Event Driven Robot Programs, Proceedings of International Conference on Recent Advances in Mechatronics (ICRAM), Istanbul, Turkey, 14-16 August, ISBN 975-518-063-X, 1995.
3. Chen, C. and Trivedi, M. T., SAVIC:A simulation, visualization and interactive control environment for mobile robots, Int. Journal of pattern recognition and artificial intelligence, Vol. 7, No. 1, pp. 123-144, 1993.
4. Chen, C., Trivedi, M. T., Bidlack, C. And Lassiter, N., An environment for simulation and animation of sensor-based robots, Applications of Artificial Intelligence IX, Vol. 1468, pp. 354-366, 1991.
5. Sick AG, Sebastian Kneipp-Straße 1, D-79183 Waldkirch, http://www.sick.de
6. Sick, Operating instructions-PLS and PLS user software laser scanner.
7. Everett, H. R., Sensors for mobile robots-theory and application, A K Peters Ltd., 1995 ISBN 1-65881-084-2
8. Brooks, A. R., A robust layered control system for a mobile robot, IEEE Journal of robotics and automation, Vol. RA-2, No. 1, March 1986
9. Moore, P. R., Pu, J., Lundgren, J-O., Ujvari, S., Intelligent Semi-Autonomous Vehicles in Materials Handling, Mechatronics conference, Skövde, Sweden, 9-11 Sept. , 1998, Elsevier
10. Martin, P., Nehmzov, U., "Programming by teaching: neural network control in the Manchester mobile robot, International conference on Intelligent Autonomous Vehicles, 1995.
11. Yen, J. Pfluger, N., A fuzzy logic based extension to Payton and Rosenblatt's command fusion method for mobile robot navigation, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 25, No. 6, pp. 971-978, June 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Modified control system for mobile robots

N. Lakota, V. Lapshov

Mechatronics and Robotics Department, Moscow Bauman State Technical University
2 Baumanskaja st. 5, Moscow, 107005, Russia

General principles of control systems creation for mobile robots are considered in this paper. The possible variants of control systems decomposition are analysed on the base of expedient distribution of the information proceeding functions and control between operator and intellectual mobile robot. Three levels structure of control system is offered. It allows to realise the different control modes for mobile robots.

1. INTRODUCTION

The development of the intelligent robots and robotic systems requires establishing of the new principles of the control systems organisation. These control systems must be adaptive and they must provide effective control of the dynamic objects under uncertain conditions.

Uncertain conditions could be defined as situations with unpredictable time-varying parameters of the system and surroundings. The main disturbances which cause in such situations are: the foreign objects appearance in working area, variation of the robot characteristics during work and so on.

The control system of adaptive robot should solve the next tasks :

- recognition of the external surroundings and its analysis;
- strategy generation of the robot's advantageous behaviour;
- creating of the necessary sequence of actions.

2. THE SPECIFICATIONS FOR THE CONTROL SYSTEM

It is possible to select the main requirements for the control system of the intelligent mobile robot.

- a system is built due to hierarchical principle and, in common case, it includes

Every of these specifications inputs the certain restriction for the control system to be created. They are formulated by the following means.

Several aims of control performance .

A robot may have several aims within one task. For example the task performed is to reach the given point in presence of obstacles which the robot has to get over. The other example is to reach a point in minimum time with energy limits. The control system has to fulfil the high priority task in presence of the interfering factors of lower priorities.

Multisensor system.

A robot usually has a lot of sensors aboard. The sensors may provide the control system with wrong measurements. However the robot must work in right way (the reliability of the data should be estimated).

Robust property.

In the case of several sensors refusal the robot has to adapt and complete its functions. It has to be capable to adequate behaviour.

Flexibility.

The control system has to provide flexible changes of the functional capabilities of the robot.

It is possible to formulate the main design principles of intelligent control systems for the mobile robots on the base of analysis the situation in area of design and creating of mobile robots for different aims:

some levels of control (strategic, tactical and servodrive), but their status can be varied;

- information supplying to the planning and making decisions processes requires the presence of information about functioning space and abilities of robot. This information must be represented in form of restoring data receiving from machine view system and from internal sensors of the robot operatively;
- information about environment must be represented in form of surrounding model (a map). It must be comfortable for as operator, as making decision system;
- a control system must provide achievement of aim with submission of possibility to make decision by the robot, beginning with real situation and with taking into account actions of the safety motion;
- an information system, that provides receiving of the current information, must be complex (integrated) and it must includes a system of technical vision, a tactile system, a navigation system, a system of determination of the robot rotary position in the surrounding and a system of internal control of the robot's state;
- a vision system has to include some channels, working on different physical principles and permitting to create a model of visible zone as separate, both in combined mode;
- an additional and very important information about the environment along the route of driving can be received on the base of analysis results for active interaction between the robot and environment, it can be saved in the form of map in memory of the onboard computer;
- operational correction mechanism of the environment current models and models of the robot abilities, including the dialogue mechanism with operator at the cases, when the reason of difference between the plan and environment can not be identified uniquely, must be presented in the control system.

3. INFORMATION SUPPLY

The equipment of the robots with technical vision means, which can determine character of surrounding at the area of robot action and can provide the motion control, is a very important problem arising when design an intellectual robots. The functional task of the machine visual system is to conduct the necessary measurements (observations), to proceed and to form models of observing scenes in borders of the given model of perception, independent from the type of machine view system. It is usual to the name descriptions formed at the output of visual system as a model of scenes or model of environment.

A number of contradictory requirements to technical devices of receiving, representation and analysis of information about environment appear when design the visual systems. For example:

- accuracy- range of action
- resolution-angle of view
- accuracy-fast action
- complexity- reliability.

Improvement of one parameter from these pairs leads to deterioration another one, as a rule.

Solving of these contradictions can be obtained on the base of unification of different visual systems with interadding characteristics to one complex (integrated) system. A principle of complexing permits to satisfy different requirements caused with variability of environment. It gives the necessary adaptation and reliability.

The environment model in local zone is a result of different device work of machine view system (MVS). The problem model is created on base of it. This model takes into consideration current ability of the robot and possibility to interact with environment.

Generalized local conception of environment model is followed from principle of complex machine view system (CMVS) creating. This conception represents the measurements results of different subsystems of machine view in form of partial local models (PLM). Action distance, view angles of subsystems and principles of information receiving are different for different subsystems. That is why degree of information reliability can be different for near and far environment zone.

The process of creating and correction of generalized local model (GLM) is represented at Fig.1.

Received data about the local zone is applied to determine obstacles and passable parts of road with taking into account possibility of the mobile robot (chassis and size of the mobile robot and so on).

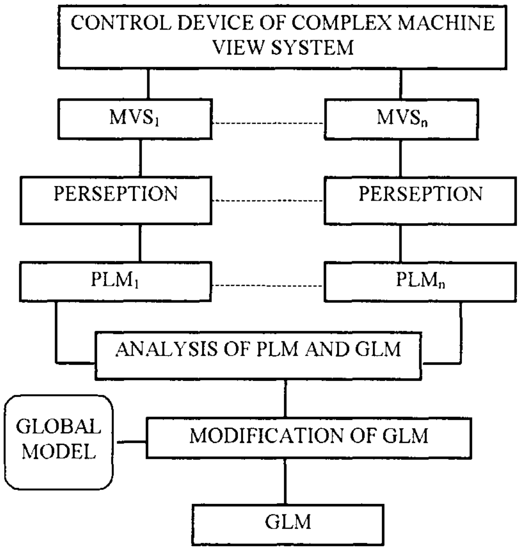

CONTROL DEVICE OF COMPLEX MACHINE VIEW SYSTEM
MVS1
MVSn
PERSEPTION
PERSEPTION
PLM1
PLMn
ANALYSIS OF PLM AND GLM
GLOBAL MODEL
MODIFICATION OF GLM
GLM

Fig 1. Scheme of creating and correction of generalized local model

4. DECOMPOSITION OF CONTROL INTO HIERARCHICAL BEHAVIOUR

The first stage of the robot designing is decomposition of the problem to the separate parts.

Decomposition of the system into the basis of internal functional modules is the usual approach in the designing. It is named "horizontal decomposition" and is presented at Fig.2.

It looks like an information loop from environment to the robot, the feedback comes through the robot action. Any changes of the separate modules (functional extension, upgrade) have to be created in the way that does not change the interfaces and functioning of another modules. The designers select the following functions usually:

- sensing,
- mapping of the sensor data in the knowledge area,
- planning,
- task performance,
- drives control [1-7].

Another approach is suggested by [8] that brings the essentially different architecture of the control system for mobile robot. The main difference is the designing of strategy for the hardware level. There appear advantages in the robust, tuning and testing properties of the CS. The basic idea of this approach is that the complicated behaviours are the composition of simple behavioural actions. The given conception is based on the external world properties and allows to avoid a traditional impasse of the artificial intellect realising when a logical model of external world is brought into the computer. In this case a vertical decomposition of the control task has place. The desirable external actions of the robot are the basis of the decomposition (See Fig.3).

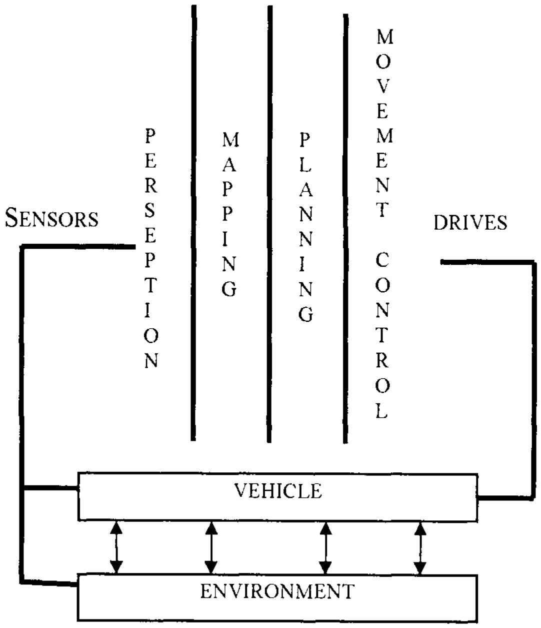

SENSORS
PERSEPTION
MAPPING
PLANNING
MOVEMENT CONTROL
DRIVES
VEHICLE
ENVIRONMENT

Fig.2 Horizontal decomposition

This architecture determines the competence levels for the mobile robot where the competence levels represent the informal specifications of the desired behaviours. Classification of the supposed competence levels:

0. Free movement without contacts with objects (external surroundings with stationary objects).
1. Search of environment.

2. Mapping and planning of the route.
3. Marking the changes at the stationary surroundings.
4. The environment's response in terms of the objects' identification.
5. Planning of the route with the environment changes considering. Reproducing of the plan into the sequence of movements.
6. Response of the objects' behaviour and modifying of the behaviour plan.

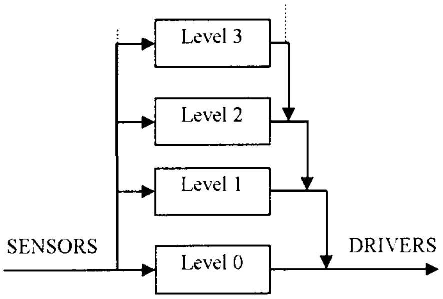


Fig.3. Vertical decomposition

Notice that every the lower competence level is included as a sublevel into the higher one. The key idea of the competence levels is that the adding of a new level to the existing set leads to receiving of the highest level.

The system designing begins from the zero level, then the first one is formed. The first level would check the zero level data, input the data into the zero level and suppress the input normal information. The zero and first levels reach together the first competency level. Ongoing this process the highest level may be achieved.

It may be shown that the structure of control system is presented in accordance with the specification of the intelligent control system listed above (several aims of the control performance, multisensor system, robust and flexibility properties).

The levels structure may be a set of processors that communicate each other by means of messages. Every processor is a complete unit that is capable to store some data inside. The processors work asynchronously, perform the monitoring of the input data and send the output messages. All the processors are designed as the similar ones in the sense of absence of the central CS. The inputs of the modules may be suppressed and the output inhibited by the output signals of another modules. This mechanism provides possibility for the higher and lower levels to work in compliance.

5. THREE LEVELS STRUCTURE OF THE CONTROL SYSTEM

The analysis of design and creating of autonomous mobile robots practice has shown that autonomous systems do not possess enough universality, flexibility and other possibilities of human intellect.

Systems with different modes of remote (with participation of human) and autonomous control are very perspective today. In this case, operators must decide more complicated tasks.

It is necessary to keep in mind that term of complexity is strong differed in artificial intelligence and human perception. It is considerably easy to automate these actions, which man have learned to do recently (for example, to read printed text or to solve differential equations) than to carry out analysis of environment with chaotic situated objects of improper geometric shape. In other hand, an automated system leads a robot up to the desirable point considerably fast, in case if coordinates of this point is given with one or another means.

The intelligent control system for the mobile robot may be represented as an hierarchical structure shown at Fig.4. Three control levels are defined in this structure - the strategic, tactical and servodrive ones.

The level structure of control system makes possible to realise different control requirements in simple way.

Connection between robot and operator, global route planning are realised at upper (strategic) level. Information basis for making decisions is provided by map data base.

Middle (tactic) level processes information received from machine view system, analyses road scenes, plans the movement trajectories in local (visible) zone.

Lower (servodrive) level realises the planning movement to the sequence of control commands for the conducting units and control the realization of it.

Regarding the semiautonomous transport vehicles an operator has chance to introduce himself into the control strategic level and to control the "macroactions" making real the supervision control (an operator on the route) or to introduce himself into the tactical level also by pointing the direction and the velocity of the movement with the help of joystick.

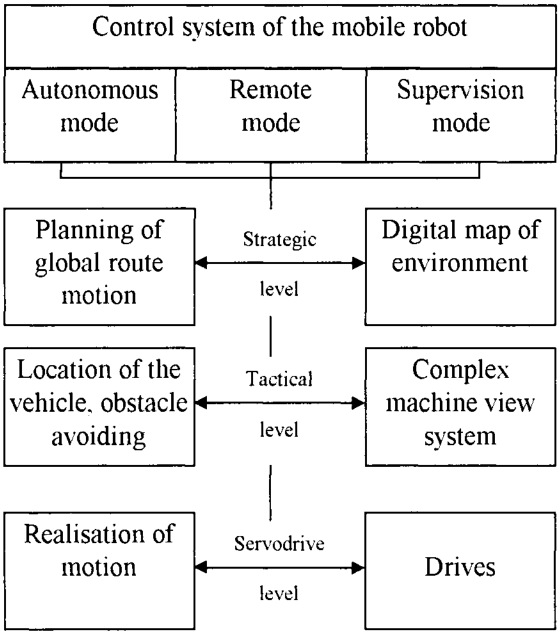

Control system of the mobile robot
Autonomous mode
Remote mode
Supervision mode
Planning of global route motion
Strategic level
Digital map of environment
Location of the vehicle, obstacle avoiding
Tactical level
Complex machine view system
Realisation of motion
Servodrive level
Drives

Fig.4. Three levels structure of the control system

In cases when the manual control is applied it is efficient to make the automatic system to find the solution for the complementary problems such as the route planning by the electronic digital map, the recognition of the robot's location at the monitor screen on this map's background, displaying the 3D model of the observing scene with its simultaneous superposition with the monotelevision displaying, the warning of the possible unusual situation to occur when a mistaken action of the operator, the registration of all the operator's action in the electronic protocol and some others.

6. CONCLUSION

Described above principles of the decomposition task for control movement of mobile robot can be applied for mobile robot design of different types. Such control systems allows to create autonomous robots or vehicles controlled by human for work in industrial environments (for example for automated assembly) and for function under dangerous conditions (for example to work at nuclear plants).

REFERENCE

1.James L. Growley, "Navigation for an intelligent mobile robot", IEEE J. Robotics Automate., vol. RA-1, '1, Mar. 1985.
2.G. Giralt, R Chatila, and M. Vaisset, " An integrated navigation and motion control system for autonomous multisensory mobile robots." in Robotics Research 1, Brady and Paul Eds. Cambridge, MA: M.I.T. 1983.
3.Y. Kanayama, " Concurrent programming of intelligent robots" in Proc, IJCAL,1983.
4.H.P. Moravec, " The stanford cart and the CMU rover", Proc IEEE vol.71, July 1983.
5.N.J. Nilson, "Shakey the robot", SRIAI Center, tech. Note 323, Apr. 1984.
6.S. Tsuji," Monitoring of a building environment by a mobile robot", in Robotics Research 2, Eds. Cambridge, MA: M.I.T.,1985
7.A. Kalyaev, J. Chernukhin, V. Noskov, I. Kalyaev, "Uniform controlling structures of adaptive robots", Moscow. Nauka, 1990.
8.R.A. Brooks,"Aspects of mobile robot visual map making", in Robotics Research 2, Eds. Cambridge, MA: M.I.T.,1984.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Dynamic Agent Architecture and Token Negotiations

Eur Ing A. A. Wallace BEng(Hons) AMIEE [a]*

[a]Open University, Milton Keynes, Bucks. England, U.K. email: AAWallace@iee.org

A great deal of work has been conducted into different architectures for agents and multi agent systems. The types of architectures range from loosely coupled to tightly coupled, some are layered and some are hybrids of types. Most of them are not very flexible or dynamic. The agents in these architectures can not reconfigure their connections to form themselves into different configurations as the environment that they find themselves in changes. In this paper a dynamic, object oriented, architecture for agents and multi-agent systems is presented along with a novel negotiation protocol.

1. INTRODUCTION

An agent in the world of AI, robotics and AL-ife is an artificial entity that does some useful work. It has some level of intelligence and is, to some degree, self-sufficient (i.e. it does not need user intervention). Robots can be considered to be real world agents and advanced robotics deals with complex systems of robots that inhabit complex, dynamic environments [1]. The agents are collected together in some way that is defined by an architecture. The architectures [2–7] that are used are, however, in general are not very dynamic. They do not reflect the changes in the agent's knowledge of the world or changes within that world. Nor do they allow the agent to have the ability to handle different varieties of tasks. In this paper an object oriented agent architecture is presented that offers agents the ability to reconfigure themselves and their interconnection with other agents to form different agents architectures. This helps the agents to handle the changes in a dynamic and unpredictable world or changes in their tasks. The architecture is built up of a number of classes that can be added to or taken away from an agent. The paper also gives a description of a number of experiments that were used to develop the architecture and a negotiation protocol [8,9].

2. AN OBJECT ORIENTED DYNAMIC AGENT ARCHITECTURE (OODAA)

The architecture presented is object oriented and, therefor, is made up of classes. It has also been engineered to be flexible. The architecture allows agents to be expanded into a multi-agent system in a variety of different ways. The architecture is also constructed to be dynamic. The agents will be able to configure themselves and then reconfigure themselves as the external environment changes. This is achieved by building the agent up out of a number of classes. These classes represent parts of the agent. The interfaces to each class are input and output communication functions which are used to send messages to each part of the agent.

Agent At the centre of the architecture is the Agent class. This is the class that collects all the other classes together and handles all the communications to other agents. In a simple system this class could be all that is needed and will then contain the intelligence of the agent, world model and anything else that it needs. In a more complex agent other classes will be needed.

Communication To help the Agent class communicate it can have any number of Communication classes. Each Communications class encapsulates a communications protocol and architecture. For example, one class could represent an Internet connection us-

---

*The work presented in this paper is towards an Open University PhD. The author would like to acknowledge the help given in preparation of this paper by Prof. G. Rzveski, Dr A. Lucas-Smith CEng FBCS of the Open University and Mrs E. Wallace BSc.

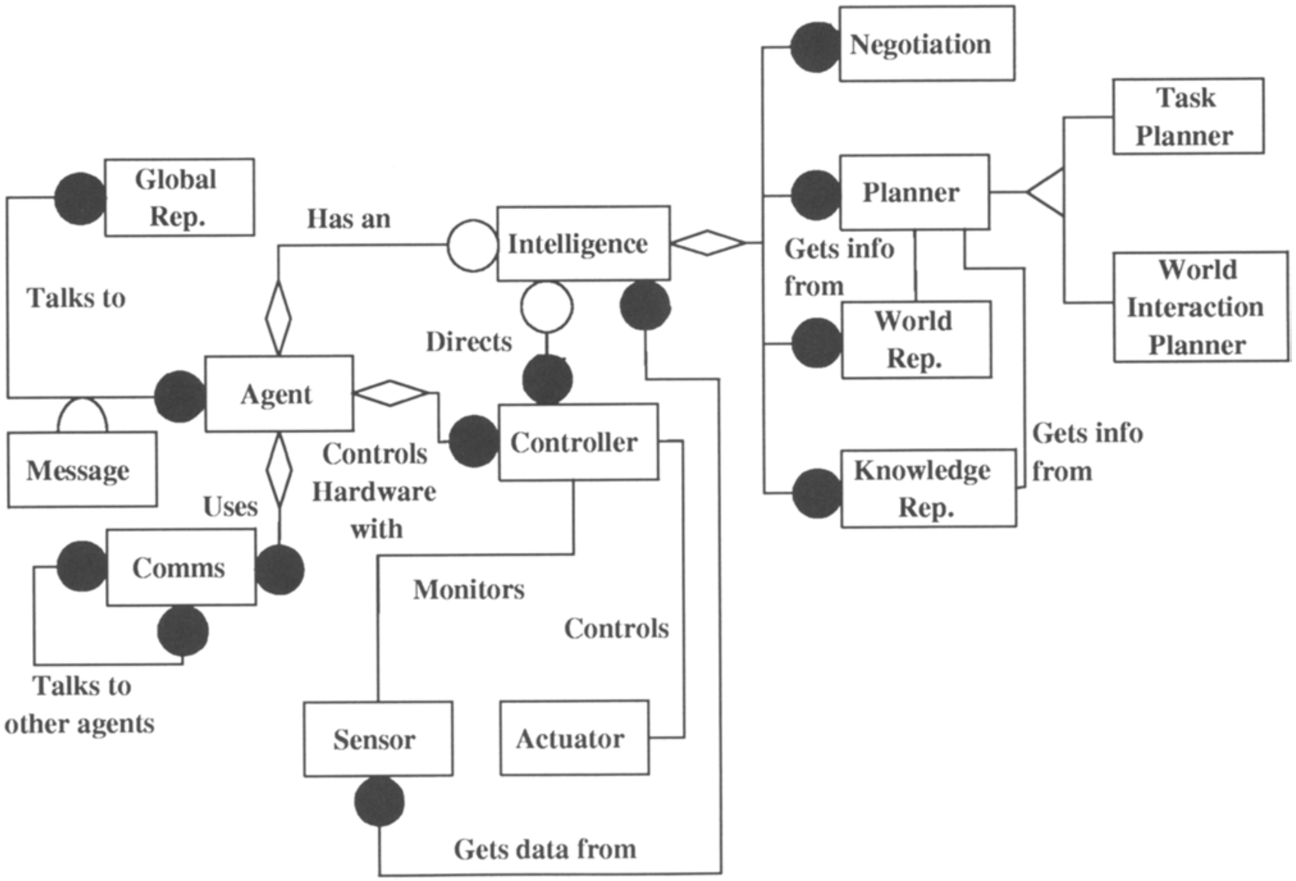

Negotiation
Task Planner
Global Rep.
Planner
Has an
Intelligence
Gets info from
World Interaction Planner
Talks to
World Rep.
Directs
Agent
Controller
Controls Hardware with
Knowledge Rep.
Gets info from
Message
Uses
Comms
Monitors
Controls
Talks to other agents
Sensor
Actuator
Gets data from

Figure 1. An Object Oriented Dynamic Agent Architecture [10]

ing TCP/IP. Another class could form an OSI seven-layer communications stack.

Intelligence The Intelligence class holds a number of classes that handle the intelligence of the agent. An agent can have at the most *one* Intelligence class.

Negotiation This class encapsulates a negotiation protocol such as the up/down token negotiation described later. This class is part of the Intelligence class. The Intelligence class can have any number of Negotiation classes.

Planner This class is a super class of two planning classes that deal with how the agent should achieve it's goals or objectives. It uses the Knowledge class to get the rules for agent.

Knowledge Representation This class holds information or knowledge that the agent needs. This could hold information such as rules for the agent interaction or how to perform a behaviour.

World Representation This class holds all the information that a robot will need about the world in which it inhabits. This would include the world model that a robot would use to navigate. This class is part of the Intelligence class. There can be any number of World Representation classes each one could be used to represent different areas or even different techniques for the same area.

Sensor The Sensor class handles the sensors of, for example, a robot. It encapsulates any type of sensor that may be used such as motor position sensors or LASER range finders.

Actuator This class represents part of the hardware that the agent uses. It gives the agent the ability to do something in the world.

Controller This class is used to control the actuator. It need not be a software item. It could, for example, be a PID controller implemented in hardware.

Message This class encapsulates the messages that are sent from class to class and from agent to agent. The Message class only holds information about the sender, receiver and the type. The actual messages are then classes that inherit from the Message class.

Global Representation This is class is global to all the agents. It holds information that all the agents need.

The Sensor class and the Actuator class are not part of the Agent class. This allows for more than one agent to have access to the actuators and sensors. This will be a source of conflict in a system with more than one agent. This conflict would have to be resolved. The resolution of the conflict could be achieved in a number of ways — the architecture does not limit the method used.

An agent is constructed of combinations of any number of these classes (with the exception of the Intelligence class where an agent can have, *at maximum*, only one). The architecture is dynamic because only the interfaces between each class are defined. This enables an agent to drop one class and replace it with another class of the same type. For example, a robot that explores an unknown environment in co-operation with other robots may require close communications with these robots. The agent could use a Communications class that allowed direct communications, in one way or another, with all the other robots that it is working with. As the environment becomes well known the robots could diverge and concentrate on their own specific tasks. This would require a more loosely coupled communications. This could be achieved by dropping the old Communication class and loading up a new Communication class that implements a different communications strategy.

3. ARCHITECTURAL EXPERIMENTS

In order to develop and to test the architecture a number of scenarios were developed.

3.1. Simple Agents

As a test bed and as a platform from which to develop the agent architecture a simple set of agents were developed. Each agent was a window with a button and a text display. The buttons were an implementation of the Sensor class and the displays were an implementation of the actuator class. Within the display a message was presented whenever an event occurred or a message was received from another agent. These agents ran as separate threads and were implemented in Java. They communicated with each other, by sending a message, whenever they detected an event. The event for these simple agents was the pressing of their button. Whenever an agent detected an event it would inform all the other agents. The agent's Communication class was directly connected to all the other agent's Communication classes.

3.2. RobotBuggy

In order to test the architecture further and develop a negotiation protocol a simple conflict scenario was devised. The scenario was made up of two simulated mobile robot buggies that were placed within an arena. The robots starting position was at one edge of the arena each robot being 90° to the other. One robot started on the left-hand side of the arena and the other at the bottom of the arena. The robots would then move towards the centre of the arena. One robot moving from left to right the other moving from top to bottom as viewed on the screen. As the robots met in the centre of the arena they were brought into conflict. Each robot attempted to continue to the other side of the arena but if they were both to do so they would collide. To prevent this collision from happening one or the other of the robots must stop and let the remaining robot pass. In order to resolve which robot is to stop they must negotiated.

For this scenario a single agent controlled each robot. The agent was rule based. Each agent directed a robot forward, maintaining track of its position in the world, and checked for the second robot by searching the local area around the robot. When the second robot was encountered the two agents negotiated. The winner of the negotiations would then continue to the far side where it would wait. The losing robot would wait until the winning robot had reached its destination before continuing. When both robots were at their respective far ends of the arena they would both continue to the start position again meeting in the middle and negotiating as before and so on.

3.3. Automatic Guided Vehicles

A more substantial test for the agents was to control the flow of a number of mobile robots in a simulated environment [11]. The mobile robots used are more commonly referred to as Automatic Guided Vehicles (AGV) which are often found in manufacturing plants. In the test scenario a number of test layouts were use. Each layout could be considered to be a lattice world model. Within the world model there are a number of points and segments. Agents were used to control the access to these points and segments. Agents were also used to control the flow of the AGVs around loops, to check their orders and to negotiate. The negotiation algorithm used was the same as in the robot buggy experiment.

Whenever an AGV was ordered to go from point to point an enquiry would be made as to weather or not there was a free path for the AGV in the layout. Enquiries were made from the AGV agent to the segment agents and the segment agents would enquire with the point agents. Enquiries would be made with other agents when necessary. This enquiry would continue until the AGV had enough segments and points to grantee free passage within the layout.

This simulation used simple agents that communicated with each other by function calls. This meant that there was no Communication class or message passing.

4. THE NEGOTIATION ALGORITHM

The negotiation algorithm is called Token Negotiation. This type of negotiation is very quick. It resolves to a winner or a loser on the first attempt. The negotiation algorithm consist of three phases as follows:

1. Token selection.
2. Token modification.
3. Token comparison.

The first phase, token selection, is to initialise each agent's token and is usually done once for each agent. The token can be initialised in a variety of different ways. The method chosen for the robot buggy agents and the AGV simulation was to randomly select a token once when the agent begins its first negotiation. Other methods for selecting the token could include assigning a token to an agent on an agent by agent basis.

The second phase is to modify the token for some criteria. Effectively the token is increased by an amount for every reason that the agent should win and decreased by an amount for every reason the agent should lose. In the robot buggy example there is no pressing reason why any one of the two agents should win over the other so no modifications were carried out on the tokens. In the AGV simulation program AGVs carrying a load would increase there token while empty AGVs would decrease there token. This gave the load carrying AGV a greater chance of winning any negotiations. In other systems there may be other reasons why one agent should win. For example, if a robot is re-charging and a second robot wishes to acquire the charger they may have to negotiate. If the second robot's batteries are almost flat and the controlling agent is almost fully charged the token may be increase or decreased by a corresponding amount. So the robot that has the charger may decrement its token as being almost fully charged is not a good reason to keep hold of the resource.

Another example is when a robot is controlled by more than one agent. If one agent is given the task of directing the agent forward and a second agent is given the task of avoiding obstacles and both agent want to control the robot at the same time. The first agent may have no reasons to alter its token but the second agent would be assigned the lowest value for a token and would increase or decease the value of that token inversely proportionally to the distance it is from an object. The closer the robot is to an object the higher the second agent's token becomes. Eventually it will gain a high enough token to win over the first agent and gain control of the robot. It would then steer the robot away from the object and decrease its token as a result. As the robot moves away from an obstacle the first agent will eventually have a high enough token to regain control of the robot and direct it once again in a straight line.

The final part of the negotiation is the comparison. One agent sends its token to a second agent. The second agent then compares the two tokens. If its token is the highest then it wins, if it is the lowest it loses. If the two tokens are the same then it can win or lose depending on how the agent has been engineered. In both the robot buggy and AGV simulation experiments the comparing agent would lose the negation if the two tokens were equal. Alternatively the token could be re-selected, modified and then the agent could re-negotiate. After negotiations the agents then modify the base token by increasing it if it lost or decreasing it if it won. This means that a losing agent has a slightly better chance of winning on the next round of negotiations. This type of negotiation is referred to as an up-down token negotiation. The next round of negotiations may take place after a time out or if new situations arises.

5. RESULTS

The first experiment with simple agents, being very satisfactorily, served to develop and refine the agent architecture. The second experiment also refined the architecture but to a lesser degree and acted as a test bed for the token negotiation algorithm. The AGV simulation gave a more substantial test of the architecture. The use of function calls was, however, unsatisfactory as it led to a non-flexible implementation of the agents but in doing so added in the development of the architecture.

As the robot buggy simulation ran the two robots would continuously met in the centre and negotiated. For the first number of tries one agent would continuously win other the other. After a time the second agent would eventually win. When that occurred the agents would then settled down into a "taking it in turns" pattern of behaviour. In the AGV simulation experiment the agents successfully control the flow of the AGVs. Within the system deadlocks were avoided.

6. CONCLUSION

The beginnings of OODAA and a number of experiments used to develop the architecture have been presented. The experiments have helped in refining OODAA but there is still a great deal of work to be done on the development of the architecture. The dynamic aspect of the architecture has not been tested at all and work will need to be done in this area. Additional work is also needed to refine the messages that may be passed from class to class as the idea of using well defined methods for each class has been shown not to be the way forward for OODAA. In the experiments most of the classes that form part of the Intelligence class have not been fully developed. More work will be needed on this part of the architecture.

Despite the fact that a great deal of work is still remaining to be done on OODAA it has, in the limited experiments so far conducted, shown a promise of the architecture being flexible and dynamic and that it could be used in dynamic and unpredictable environments.

One part that has been demonstrated to work well in both the robot buggy and the AGV simulation experiments is the up-down token negotiation algorithm. This is a sub type of the more general form of a token negotiation. Some more work would be needed in testing this in other situations. Additionally, more work could be done to test different forms of token negotiation.

7. FUTURE DEVELOPMENTS

One future intention of the agent architecture is to build it up into an agent engineering system. In this system agents would not be programmed but instead they would be constructed within a development environment. An agent would be assembled from predefined classes and tested within the environment. A specification would then result from this. A real agent, to construct itself out of the specified classes, would use this specification.

REFERENCES

1. J. O. Gray, "Recent developments in advanced robotics and intelligent systems," *Computing and Control Engineering Journal*, vol. 7, pp. 267–276, December 1996.
2. F. R. Noreils, "An architecture for cooperative and autonmous mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation, Nice, France*, pp. 2703–2710, Alcatel Alsthom Recherche, 1992. email noreil@aar.alcatel-alsthom.fr.
3. R. Kazman, "Hidra: An architecture for highly dynamic physically based multiagent simulations," *IntJCompSim*, pp. 149–164, 1995.
4. B. Kipper, "Interlocking multi-agent and blackboard architectures," in *Progress in Artificial Intelligence*, pp. 371–375, Oct 1995. email kipper@cs.uni-sb.de.
5. P. R. Cohen, A. Cheyer, M. Wang, and S. C. Baeg, "An open agent architecture," in *AAAI Spring Symposium*, 1994.
6. J. S. Zelek, *SPOTT: A Real-Time, Distributed and Scalable Architecture for Autonomous Mobile Robot Control.* PhD thesis, Centre for Intelligent Machines, Dept. of Electrical Engineering, McGill University, 1996.
7. R. Goodwin, K. Z. Haigh, S. Koenig, and J. O'Sullivan, "A layered architecture for office delivery robots," in *First International Conference on Autonomous Agents*, pp. 245–252, February 1997.
8. S. Kraus, J. Wilkenfeld, and G. Zlotkin, "Multiagent negotiation under time constraints," *Artificial Intelligence*, vol. 75, pp. 295–345, June 1995.
9. S. Parsons and N. R. Jennings, "Negotiation through argumentation — a perliminary report," in *Proceedings of the International Conference on Multi-Agent Systems, Kyoto.*, 1996.
10. J. Raubaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen, *Object-Oriented Modeling and Design.* Prentice Hall International, 1991.
11. A. Wallace, "Flow control of mobile robots using agents," in *Proceedings of the 29th International Symposium on Robotics*, 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Automatic navigation in a complex diffusion field environment

E. E. Kadar* and G. S. Virk†

* Department of Psychology
University of Portsmouth, Portsmouth, Hampshire PO1 2DY, UK.

† Department of Electrical and Electronic Engineering
University of Portsmouth, Portsmouth, Hampshire PO1 3DJ, UK.

The paper discusses the problem of target searching strategies for autonomous agents operating in a complex chemical gradient field. In previous studies the authors argued that chemotaxis is not a feasible navigational strategy in natural settings because it is highly sensitive to environmental perturbations. Computer simulations have demonstrated that in naturally unstable diffusion fields the seemingly inefficient biased random walk is a better strategy for both static or moving targets. The present study further generalises these findings to a complex diffusion field generated by multiple targets which are sources of diffusive chemicals.

1. INTRODUCTION

As autonomy plays a larger part in mobile robotic systems the role of navigation becomes more important so that the machine can find its way around the working environment and carry out its tasks in an effective manner (Maes [1]). There are various methods of achieving satisfactory solutions; the simplest ones use a basic feature-following philosophy to find their way in well-defined situations (Zelinsky [2]). When the environment becomes less stable such as that encountered in realistic situations the difficulty of the problem increases by several orders of magnitude and, in some cases, becomes totally unmanageable using current techniques. These methods are mostly based on a representation of the working environment and often use a set of simple rules.

In an unknown or dynamically changing environment, these approaches have several shortcomings. These include the need to regularly update the representation to avoid collisions and getting lost, and because the different navigational rules have different efficiencies in different environments, dynamic rule selection is usually necessary. All this leads to very demanding computational requirements to achieve real-time performances. Most importantly, the rule-based approaches lack the much needed flexibility common in animal navigation. Flexibility is intrinsic to goal-directed biological systems "designed" by evolution to cope with the intrinsic variability and uncertainty in natural settings. Robot designers would welcome similar flexibility in their solutions. However, researchers typically fail in their efforts because they continue to seek optimality at the expense of flexibility in their solutions.

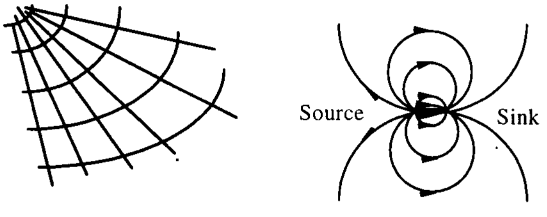
Using field theory provides a promising alternative strategy. Kadar [3], for instance, has outlined a generic field theoretical approach to navigation which captures the remarkable plasticity of animal locomotion in a chemical diffusion field. He has questioned the common belief that animals use chemotaxis in finding a source of a chemical diffusion field. Chemotaxis is a mechanism to generate a smooth movement trajectory ideally connecting an initial position to the source of the diffusion field. Although this method seems plausible its applicability is severely limited because such diffusion fields are seldom stable in natural environments.

To further clarify these points, the paper starts by presenting an overview of the theoretical principles of field based navigational strategies. This is followed by a discussion on navigation in diffusion fields. Section 4 then presents a summary of previous results obtained by the authors via computer simulations (Kadar and Virk [4], Virk and

Kadar [5]); these results demonstrate the superiority of the biased random walking in comparison to a strategy based upon chemotaxis. Section 5 then presents new results for more complex diffusion field environments with multiple sources as targets. The results presented show that the biased random walking strategy is a flexible approach for target searching in these conditions.

2. THE CONJUGATE FIELD APPROACH

In robot navigation, field theoretical methods have already been used (see for example Krogh [6]; Payton [7]). Most of these attempt to design an optimal movement trajectory from an initial position to a target location by using the representation of the environmental layout. Such representational approaches, however, have several major shortcomings; for instance, they require that the environment layout be either fed into the robot in advance or a cognitive map be learned by an exploratory navigation process. Furthermore, representational approaches are very sensitive to changes in the layout forcing robot navigators to include complex on-line instructors or built-in heuristics to continuously revise and update their inner models.



(a) Gradient field (b) Bipolar source-sink

Figure 1: Gradient fields for navigation

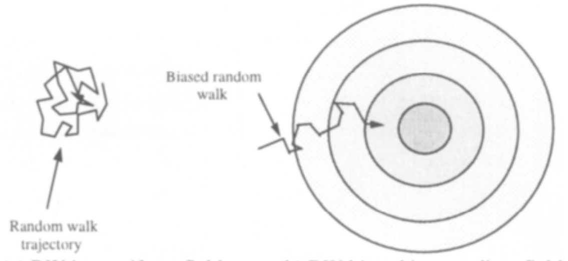
Humans and animals need to navigate mostly in an unknown or continuously changing environment. Thus, the much needed flexibility probably forces them to use non-representational strategies. Despite the variety of perceptual modalities used by animals, Kadar [3] has shown that navigation in a cluttered environment can be described by a core of fundamental field equations. This core field can be construed from the information field, viewed as a gradient field, and the commonly used "taxis" behaviour of animals. It is well known, that animals have developed various mechanisms (phototaxis, geotaxis, etc.) by which they can move towards a target using some form of external stimulus (a light source, centre of gravity, etc.). The various forms of taxis behaviour are basically embodiments of the gradient technique commonly used in mathematics and applied to physical problems. A common feature in these techniques is that the trajectories of motion are orthogonal (conjugate) to specific fields (see Figure 1(a)).

During the navigation of an animal the concentration levels of the diffusion field (of a chemical substance) can be viewed as the information field. It is commonly believed that navigation towards a source of chemical substance is possibly guided by measurement of the concentration value and information fields as the gradient fields used to generate orthogonal conjugate fields of all possible motion trajectories. These pairs of fields can be described by a common Laplacian differential equation with the proper boundary conditions (e.g., $Lv = 0$). If the sources of an information field are included in the domain of motion then Poisson equations (for example $Lv = S$) are needed. Kadar [3] revised these equations in the context of animal navigation by reinterpreting the information sources as navigational sinks (the goals to be reached) and introducing the initial positions of the animals as navigational sources. In this way a simple ballistic motion can be described as a bipolar field of a source-sink pair as shown in Figure 1(b).

These field equations can be used to capture navigational strategies by various fundamental perceptual modalities (such as vision, hearing, and mechanical vibration). Although it is common to assume that chemotaxis is used for navigation in a chemical diffusion field, the actual behaviour of dogs, cats, and other animals with a good sense of smell clearly suggest a different strategy. These animals exhibit a strange "sniffing" behaviour. Instead of moving straight to a "smelly" target, they move slowly and often make turns while moving their heads around. Despite this meandering movement they gradually drift toward the target and finally reach it. The explanation of this behaviour can be found in the fundamental properties of the diffusion process.

3. NAVIGATION IN A DIFFUSION FIELD

Smell and taste are the most fundamental senses because they are directly related to the vital metabolic processing. Higher-order species use smell for various other purposes. Smell can be used to identify a mating partner, find food, detect predators, consider scent marked territories, etc. Most importantly, various animals including the lowest level in the hierarchy of species, bacteria use chemical substances and their associated gradient fields as aids to navigate. Although there are numerous experimental results on the perceptual control of navigation (Kleerekoper [8]; Papi [9]) the systematic investigation of these processes in higher-order species poses great difficulties due to the complexity of their behaviour (e.g., several interacting perceptual modalities, high degrees of freedom in movement control). These complex issues have led researchers to investigate simpler creatures such as bacteria's perceptual and control processes in response to chemical gradients (Koshland [10]; Lackie [11])



(a) RW in a uniform field (b) RW biased in a gradient field

Figure 2: Biased random walk

It is argued that the search behaviour in a diffusion field is different than in other fundamental (vibratory, electro-magnetic) fields because the diffusion equation has a first-order temporal partial derivative. Field equations provide macroscopic descriptions of the motion of elementary particles. Diffusion processes, however, can also be described as microscopic phenomena and diffusion is a random motion of small particles arising from thermal energy. Since Einstein's classic paper in 1905, it is known that at absolute temperature, T, the average kinetic energy associated with movement in any direction equals kT/2, where k is Boltzman's constant; this formula is valid independently of the size of the particles. Particles that are visible under the microscope exhibit Brownian motion (or otherwise known as a random walk process), which is discussed thoroughly in basic textbooks of stochastic processes. The main features of this process are as follows:

(a) A particle tends to explore a given region of space and tends to return to the same place several times before it drifts away to explore another region thoroughly;
(b) It has no hysteresis effects (memory) and may drift back to a region already explored;
(c) the "drifting away" processes, if present, are usually slow;
(d) External forces (e.g., wind, current) can change the random walk processes and they can be described by additional drift forces.

Whilst random walk processes are characteristic of inanimate particles making them look animate, random walking, has been observed in micro-organisms making them appear similar to inanimate diffusive particles. This virtual similarity, however, hides the fundamental difference between these two processes. The Brownian motion of inanimate particles is due to thermal forces in contrast with the self-propelled motion of bacteria.

Berg and Brown [13] used a motion tracking microscope to observe Escherichia coli (E. coli) bacteria. To simplify the analysis the normally 3D motion can be constrained to 2D by using a thin layer of fluid in a dish. Under these conditions, the movement trajectory is shaped by two different modes of motion, namely running straight ahead (translation) or tumbling (rotation). Bacteria are seen to switch between these two fundamental modes of motion creating zigzag shaped trajectories. In a homogeneous medium the motion is an aimless random walk process. In the presence of a gradient field of a chemical substance the random walk process is modified by a drift which is a result of internal processes. If the length of the straight movement segments (the steps) is dependent on the concentration of the medium, an overall drift can be achieved. To be more precise, if the run length is in correlation with the gradient of the medium, that is the direction is biased by the chemical gradient, the bacteria gradually move toward the source of a nutritious substance or move away from the source of a noxious substance.

At a first glance this technique may not be the most efficient since there are lots of short detours. However, there are good reasons for using the biased random walk strategy for navigation. First of all the intrinsic instability of a naturally slowly changing diffusion field can make the chemotaxis useless. Chemical gradients are usually weak and easily disturbed by other processes (e.g., wind, fluid current) and so to use such a gradient for orientation towards the source, animals must be able to obtain a valid solution in the presence of perturbations. Also, for chemotaxis two distinct but simultaneous measurements are needed, and in a weak field, the two measurements (sensors) have to be sufficiently far from each other for achieving the desired resolution in measurement, so that an appropriate decision on the direction of steering can be made.

Perception of chemicals and orientation via olfaction play important roles in higher-order species as well (e.g., insects, birds, fish, etc.). Previous research projects on the use of chemical gradients for navigation do not provide clear-cut evidence to support a common mechanism for perceptual control because the behaviour of higher-order species is much more complex. Nevertheless, from the arguments presented it is plausible to believe that higher-order species also use the same types of strategies. For example, dogs and cats slowly move while turning their heads around when they are sniffing to find food. In the light of the proposed theory these obvious observations can be experimentally investigated in the future, but computer simulation can also be used providing an easy way to control experimental conditions. The following case studies have already tested and validated the proposed theory by computer simulations (Kadar and Virk [4]; Virk and Kadar [5]).

4. SUMMARY OF PREVIOUS CASE STUDIES

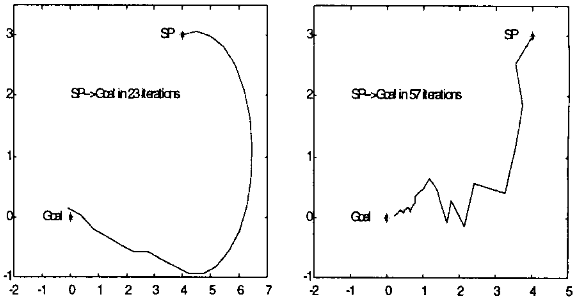
Kadar and Virk [4] have demonstrated the key points of the proposed field theory by computer simulations of navigation. To test the superiority of the biased random walk strategy over chemotaxis they have used different conditions in stable and noisy diffusion fields. Specifically, the navigational conditions were as follows:

1) chemotaxis in a stable diffusion field;
2) chemotaxis in a noisy (unstable) diffusion field;
3) biased random walking in a stable diffusion field; and
4) biased random walking in a noisy (unstable) diffusion field.

In each condition the same starting position (SP) and a static goal (G) as a source of the diffusion field of a chemical attractant has been used.

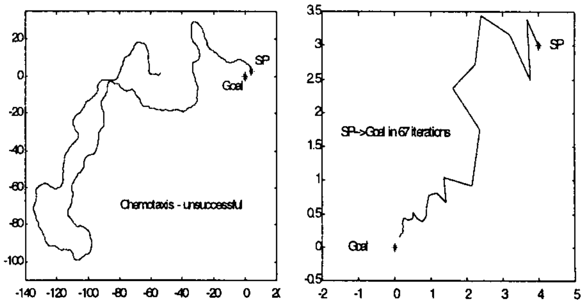
During chemotaxis the rotation towards the target in each step is either 0 or 0.25 radians. It is small relative to the step size of 0.5 and this slow steering explains the curvilinear trajectory in Figure 3(a). A uniform random number generator R(-0.5,0.5) was used for both the direction selection in the random walk and in the noise creation to perturb the diffusion fields.

Figures 3 and 4 provide a sample trial under each of these situations supporting the hypothesis; chemotaxis is seen to be optimal under stable conditions except for the slow steering mentioned above. Figure 4(a) shows that the chance of finding the source in a noisy diffusion field using chemotaxis is rather slim and most of the trials carried out were successful - that is, using 1000 steps per trial were not usually sufficient to find the goal. In contrast, the biased random walking strategy was successful under both stable and noisy field conditions (see Figures 3(b) and 4(b).



(a) Chemotaxis (b) Biased random walking

Figure 3: Comparison of strategies in a stable field



(a) Chemotaxis (b) Biased random walking

Figure 4: Comparison of strategies in a noisy field

These simulations have demonstrated that the biased random walk approach can form the basis for a robust and flexible navigational strategy. In general, chemotaxis is inferior to biased random walking even though under specific conditions, such as those found under stable diffusion fields and/or high concentration levels, it can also be used with success. Virk and Kadar [5] have also shown that a biased random walk is a robust strategy in catching moving targets.

5. SEARCHING IN COMPLEX FIELDS

Having tested the method with one target, the next step is to apply the random walking process to situations where several targets need to be visited. In nature, bumblebees solve a similar task during their feeding behaviour. Pyke [14] observed that bumblebees in the Colorado Mountains gathered nectar from flowers in a surprisingly efficient way. The flowers were clustered and varied considerably in terms of their nectar content. Pyke studied individual marked bumblebees while they were visiting flowers not visited before by other insects. It was noticed that bumblebees returned to previously visited flowers only 4 times out of 482 observations. Researchers (see for example Griffin [15]) are puzzled by this amazing performance and speculation continues whether the insects have special memories or use special scent marking to avoid previously visited flowers. None of these strategies seems plausible in reality. Because these insects have compound eyes it is often believed that they use vision to develop a cognitive map. Although insects have pattern recognition capabilities it is unlikely that these are complex in an significant way. Using scent marking seems a more plausible strategy, but is unnecessary because bumblebees can simply use the intensity of the nectar diffusion fields to discriminate individual flowers. The visited flowers become depleted sources of diffusion fields and hence they are avoided later. It is obvious that bumblebees use vision and chemical sensing in a co-ordinated manner when flowers are being visited. Vision makes the behaviour more straightforward but the use of chemical sensing is also crucial.

To test the potential use of chemical sensors in the process of biased random walking in a complex diffusion field, a simulated environment was used based upon the superposition of the diffusion fields of three targets (T1-T3) on a two dimensional layout. The diffusion fields of the three targets were assumed to be of equal strength but when a particular target is reached its diffusion field is switched off. Figures 5 and 6 show sample results of successful trials for this difficult search scenario when the starting position of the agent is approximately equidistant to all the three targets. The figures show that in different trials the order in which the targets are reached is random due to the symmetric nature of the simulated conditions. In contrast, we can consider non-symmetric conditions where the starting position of the agent is close to one of the targets. Figure 7 shows three trials for the case when the starting position of the agent is close to Target 1; all the three runs resulted in Target 1 being reached first.

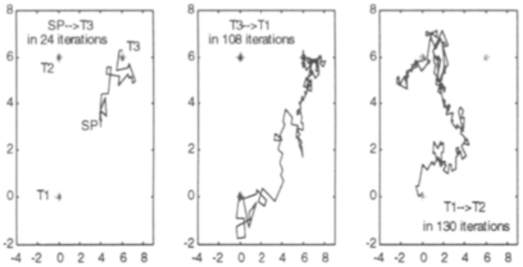


Figure 5: Sample trial 1 with central SP

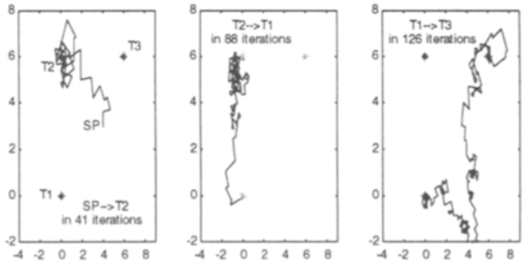


Figure 6: Sample trial 2 with central SP

6. CONCLUSIONS

The paper has investigated the potential use of a new field theoretic approach for autonomous navigation proposed by Kadar [3]. In particular, the optimality and plasticity of navigation in a complex chemical diffusion field has been discussed. The proposed conjugate field approach suggested in previous studies Kadar and Virk [4] indicates that a biased random walk is superior to chemotaxis (in terms of flexibility and robustness). In this paper computer simulations have clearly demonstrated that a biased random walk is a flexible strategy in seeking the sources within a complex chemical diffusion field. These findings are encouraging in designing robot navigators for possible industrial applications, where chemical signals can be used in defining the field.

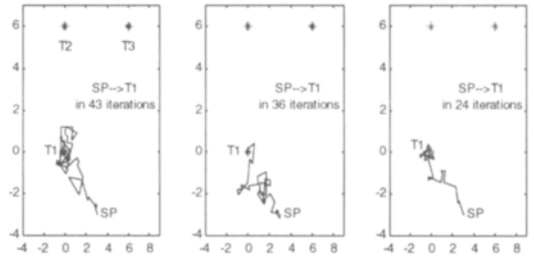


Figure 7: Segments of trials with SP near Target 1

Similar situations are encountered in industry where several machines need to be supplied with material for processing and this transportation can be carried out by a mobile robot. Traditionally, such tasks have been carried out by a machine with some fixed rule-based mechanism optimised in some manner. However, in practice more flexibility would be an advantage since the depletion rate of each machine is variable, and hence the replenishment and navigational task planning for the robot can be highly complex. A field theory approach could have merit in such applications so that when a particular machine is running low in material it can emit a chemical substance thereby requesting attention from the robot. In this way the mobile robot can establish in advance whether the machine requires servicing or not and hence plan its long- and short-term navigational paths appropriately.

REFERENCES

[1] Maes, P. Designing autonomous agents: Theory and practice from biology to engineering and back. Branford Books. MIT Press, 1990.

[2] Zelinsky, A. A mobile robot exploration algorithm: Robotics and Automation, Vol. 8, No 6, December 1992, pp 707-717, 1992.

[3] Kadar. E. E. A field theoretic approach to the perceptual control of action. Unpublished Dissertation. University of Connecticut, 1996.

[4] Kadar, E. E. and Virk, G. S. Field theory based navigation for autonomous mobile machines, Proceedings of the IFAC Workshop on Intelligent Components for Vehicles, 23-24 March 1998, Seville, Spain, 1998.

[5] Virk, G. S. and Kadar, E. E. Field theory based navigation towards a moving target, Proceedings of the 29th Int. Symposium on Robotics: Advanced Robotics - Beyond 2000, 27-30 April 1998, Birmingham, UK, 1998.

[6] Krogh, B. H. A generalised potential field approach to obstacle avoidance control. International Robotics Research Conference. Bethlehem. PA. August, 1984.

[7] Payton, D. W. Internal plans: A representation for action resources. In Maes P. (ED.). Designing autonomous agents: Theory and practice from biology to engineering and back. pp. 89-103. Brandford Books. M.I.T. Press, 1990.

[8] Kleerekoper, H. Olfaction in fishes. Bloomington, IN: Indiana University Press, 1969.

[9] Papi, F. Animal homing. New York: Chapman and Hall, 1992.

[10] Koshland, D. E. Jr. Bacterial chemotaxis as a model behavioural system. New York, NY: Raven Press, 1980.

[11] Lackie, J. M. Cell movement and cell behaviour. London: Allen and Unwin, 1986.

[12] Berg, H. C., and Brown, D. A. Chemotaxis in Escherichia coli analysed by three-dimensional tracking. Nature, 239, 500-504, 1972.

[13] Pyke, G. H. Optimal foraging in bumblebees: Rule of movement between flowers within inflorescences. Animal Behaviour, 27, 1167-1181, 1979.

[14] Griffin, D. R. Animal thinking. Harvard University Press. MA: Cambridge, 1984.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Artificial evolution of street-crossing robots

M. Wahde[a]

[a]NORDITA, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Simulated robots capable of crossing several rows of moving obstacles are designed using artificial evolution. The simulation method is described and the neural network controllers of the robots are analysed and discussed. Some important issues for further development are introduced and briefly discussed.

1. Introduction

The construction of machines with the ability to carry out everyday tasks, such as moving (with a minimum of collisions) through a crowd of people (or other obstacles) or crossing a busy street, constitutes an important problem in robotics. A robot capable of navigating through a crowd would be useful in several applications, e.g. as an autonomous delivery or transport system.

The ability of humans and other animals to carry out everyday tasks is the result of very long periods of evolution, and normally one does not have to think about *how* one crosses a street or moves through a crowd. Therefore, it is difficult to design – in the usual sense – robots capable of exhibiting such behaviour.

Thus, while traditional AI techniques are being applied with increasing success to the problems of robot motion planning and navigation as well as image recognition and understanding, it is of interest to investigate how evolutionary methods (such as genetic algorithms) can cope with these problems.

The advantages of using evolutionary techniques are manifold. First, it is possible, using such methods, to simultaneously optimize both the architecture and the parameters of the robot control system [1,2]. Second the amount of human-introduced bias is minimal. Third, it is possible to evolve solutions which employ counter-intuitive tactics to solve the problem and therefore would be very difficult to design by hand (see [3], for an example of several evolved strategies for locomotion). Finally, evolutionary methods have the aesthetic advantage of using procedures which at least resemble those by which natural control systems, i.e. brains, evolved.

However, there are also some disadvantages. For example, the evolutionary process often involves the very time consuming evaluation of a large number of individual robots. In addition, the evaluation time of individual robots may scale disfavorably with the size of their control systems, making it difficult to evolve systems of sufficient size to be applicable to realistic situations.

Nevertheless, evolutionary methods can yield interesting solutions to robot navigation problems, and are certainly worth exploring. In section 2, the simulation method is described in detail. Section 3 contains a general description of the simulations. Some preliminary results of the investigation, which is still in progress, are presented in section 4. The results are discussed in section 5, where also the conclusions are presented.

2. Method

In order to investigate the suitability of genetic algorithms for evolving simulated street-crossing robots, a modified version of a computer code written for artificial life applications [1] has been used. In this code, the simulated robots consist of a circular body and a continuous–time recurrent neural network "brain", which receives visual inputs and produces motor output signals.

The neural networks are obtained from a growth process, the parameters of which are decoded from a chromosome with 13 genes (which take values in the range 0-9) per neuron: gene 1 codes for the (horizontal) layer in which the

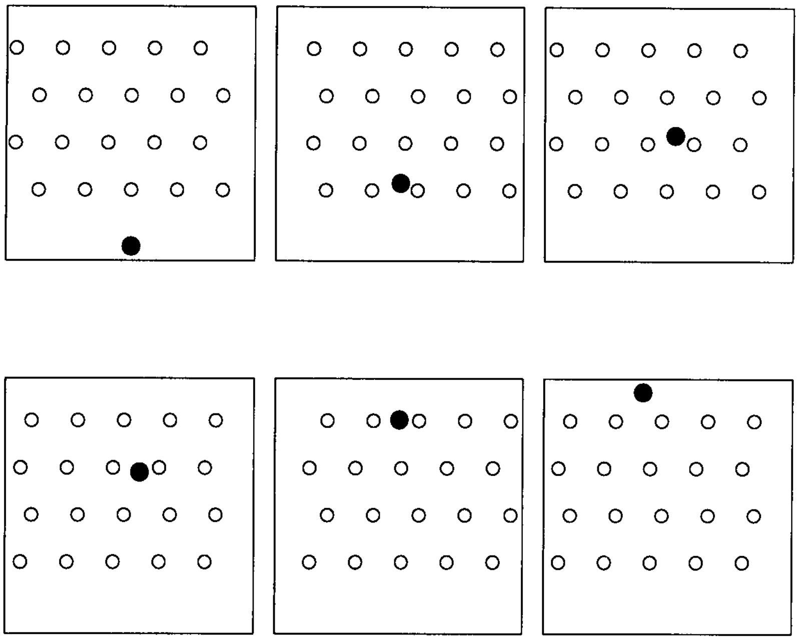


Figure 1. Images from an orbit of an evolved robot (filled disc), shown at crucial moments when RoO (circles) were being crossed. The upper left panel shows the starting configuration.

neuron is situated, and genes 2–4 determine its horizontal position within that layer.

For visual input neurons, genes 5 and 6 determine the visual input angle, i.e. the circle sector in which the neuron obtains visual inputs. The input signal equals the fraction of the circle sector which is occupied by obstacles. The positions on the robot of the visual neurons are obtained by mapping the horizontal positions to the unit circle. Thus, the visual input neurons can point in any direction.

For internal neurons, genes 5 and 6 encode the direction of neural growth (which is vertical for input neurons). Genes 7 and 8 encode the opening angle of the neural growth cone, i.e. the circle sector in which the neuron connects to other neurons. The neural growth cone is centered on the direction of neural growth, as obtained from genes 5 and 6. Finally the connection weight at distance r from the neuron is given by $w = w_0 e^{-r/l_{\text{scale}}}$, where w_0 and l_{scale} are determined by genes 9–11 and 12–13, respectively. Arbitrary connections are allowed between the neurons, so that internal circuits can form in the network.

The neurons of the network are described by the equation

$$\tau \frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + \sigma\left(\sum_j w_{i,j} x_j\right) + I_i, \qquad (1)$$

where τ is a time constant, x_i denotes the activation level of neuron i, and I_i is the visual input (only relevant for input neurons). σ is the neuron activation function, which takes values between 0 and 1. The equations of motion of the robots are

$$m \frac{\mathrm{d}v}{\mathrm{d}t} + c_l v = k_l \frac{(M_{\mathrm{l}} + M_{\mathrm{r}})}{2}, \qquad (2)$$

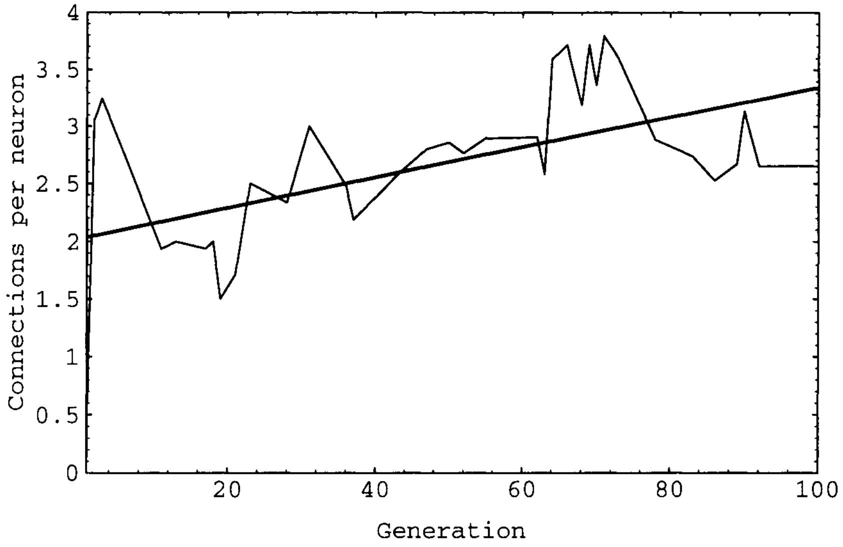


Figure 2. The network complexity, measured as the number of connections per neuron, as a function of generation for a successful run. As shown by the least-squares fit (straight line), the complexity increases from around 2 in the early generations to around 3 in the later generations.

$$\frac{\mathrm{d}s}{\mathrm{d}t} = v, \tag{3}$$

$$I\frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} + c_a\dot{\theta} = k_a\frac{(M_\mathrm{l} - M_\mathrm{r})}{2}, \tag{4}$$

and

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \dot{\theta}, \tag{5}$$

where m, c_l, k_l, I, c_a, and k_a are constants specified in the input data, and M_l and M_r are the left and right motor signals, respectively. s denotes the curvilinear coordinate along the direction of motion of the robots, which is denoted by θ.

3. The simulations

The world in which the robots operated consisted of a rectangle (with periodic boundary conditions in the horizontal direction) containing horizontally moving rows of obstacles (hereafter RoO), simulating a busy street. The robots started in the position indicated in the upper left panel of Figure 1, and their goal was to navigate, in a given time, past as many RoO as possible and preferably reach the top of the rectangle. The vertical positions and radii of the obstacles as well as the (constant) speed and direction of each of the RoO were specified in the input data.

In each simulation, a population of typically 30–100 robots was generated by assigning random numbers to the genes of the chromosomes. For each robot, the neural network was obtained by decoding the chromosome, and then the performance of the robot was evaluated by allowing the robot to move for 10,000 time steps.

The fitness measure ought to be such that rapid movement in the vertical direction is rewarded whereas collisions are discouraged. There exist several fitness measures meeting these requirements, for instance

$$f = \Delta y - \sum_k p_k, \tag{6}$$

or

$$f = \frac{\Delta y}{(1 + \sum_k p_k)}, \tag{7}$$

where f is the fitness, $\Delta y (\equiv y_\text{final} - y_\text{start})$ is the distance travelled in the vertical direction,

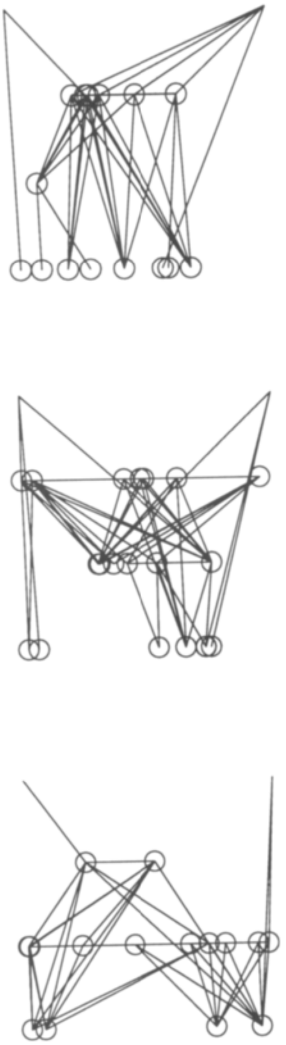


Figure 3. The neural networks of the best robots from three different runs. The lowest row in each panel shows the input neurons, which are mapped to the circular body of the robot. The left and right motor connections of each network are shown at the top of each panel. The direction of signal flow is *not* indicated in the figure.

and $p_k(= p = \text{constant})$ is the penalty for collision k. When all robots in the population had been evaluated, and fitness values had been assigned, the next generation was formed using a genetic algorithm with fitness-proportional selection and mutations. Several different values of the crossover probability (p_{cross}) and the bitwise mutation probability (p_{mut}) were tested. The results were only weakly dependent on the values of these parameters, and in the simulations, they were kept fixed at $p_{\text{cross}} = 0.7$ and $p_{\text{mut}} = 0.015$.

An important issue is that of *generalization*. If all the robots were tested against one and the same configuration of obstacles, then, clearly, successful robots would be those that adapted to the particular configuration rather than learning to handle general situations. In order to obtain robots more likely of being capable of generalization, each robot in (most of) the simulations described below was tested against three different configurations (with different speeds and positions of the obstacles), and the fitness was taken as $f = \min(f_1, f_2, f_3)$, where f_1, f_2, and f_3 denote the fitnesses obtained for each obstacle configuration. Thus, for each robot evaluation, the equations of motion were integrated for a total of 30,000 time steps.

4. Results

A number of simulations were carried out, with different parameter settings for the genetic algorithm and the obstacles. The input parameters determining the equations of motion for the robots were kept fixed at $m = 1.0$, $c_l = 0.1$, $k_l = 0.02$, $I = 0.5$, $c_a = 0.15$, and $k_a = 0.012$. In the early simulations, all the RoO moved in the same direction and with the same speed (and only one configuration of obstacles was used in the evaluation). Later simulations involved RoO moving in different directions and with different speeds, and the robots were tested against three configurations of obstacles as described above. It was found that, in most cases, the artificial evolutionary process produced robots capable of crossing several RoO.

In all simulations, the robots were started with their forward directions facing the obstacles.

Several different fitness measures were tested, including the ones defined in Equations (6) and (7), and the results were found to be rather insensitive to the particular measure used. After these tests, the measure defined in Equation (7) was used in most simulations.

The level of the collision penalty determined the behaviour of the evolved robots. For small p (< 0.1), the robots reached far in the vertical direction, but with a number of collisions. For large p (> 0.5), the robots only traversed one or two RoO, and often moved along the RoO, as if looking for an opening where a safe crossing could be made.

The orbit of an evolved robot is shown in Figure 1. However, since both the robot and the obstacles were moving, it is difficult to represent the orbits in non-moving images. In order to better illustrate the orbits, several animations were made. Some of these are available on the WWW at http://www.nordita.dk/~wahde/sc.html

4.1. Properties of the evolved networks

Starting from the initial, random networks, evolution rather quickly produced successful robots - typically, the fitness values reached 50 % of their final values after only a few generations. Furthermore, as shown in Figure 2, the network complexity, measured by the number of connections per neuron, tended to increase during the simulations. An analysis of the best networks from each generation indicated that the evolutionary process first attempted to find simple networks that could provide a crude solution to the problem, and then refined those networks by adding more connections between the neurons.

The number of neurons used in the networks stayed almost constant during the simulations. However, there was a weak trend towards an increasing number of backward connections. Such connections also enable more complex behaviour, by delaying or reinforcing signals present in the network.

4.2. Generalization

In order to investigate the general characteristics of the robots, and to examine to what extent they were able to handle general situations, the best robot obtained in a long run (100 generations) was tested against several configurations of obstacles, most of which contained more RoO than were used in the simulation that generated the neural network of the robot. In the tests, the robot was allowed to move in a single configuration of obstacles for as many as 20,000 to 50,000 time steps.

The results were mixed: In general, the robots could handle rather well configurations in which the RoO moved with approximately the same speed as in the run which generated the neural network. However, in cases where the RoO were moving either slower or faster, even by rather small amounts, the robots failed to pass them without colliding with at least one obstacle, and sometimes failed altogether and remained stuck behind the lowest row.

5. Discussion and conclusion

The simulated robots described above were generally able to satisfactorily solve the navigation problem, despite the very large size of the search space (10^{390}, in cases with 30 neurons). In such large search spaces, there presumably exist many different neural networks which allow the robots to function properly, much as natural evolution has produced a variety of creatures which exhibit similar behaviour. This hypothesis is supported by the fact that the neural networks of the best robots from different runs had *very* different appearance, as shown in Figure 3. Thus, one should not expect unique solutions to the type of problems dealt with in this paper.

As mentioned above, the networks only showed limited capability of generalization. There is a number of ways of obtaining better networks. For instance, other values of the parameters for the equations of motion could be explored, the simulations could be run for a longer time, and each robot could be tested against a larger set of different obstacle configurations.

The purpose of the present investigation has been to increase the understanding of basic features of artificially evolved neural network controllers, rather than attempting to realistically model the situations real robots would face.

Thus, the setup for the simulation was a fairly simple one, and the signals to and in the network were taken to be noise-free.

Furthermore, the robots themselves were rather simple, such that a visual input was required to obtain useful motor output. More realistic simulations, aimed at producing control systems for actual, real-world robots, would have to take into account the noise in the environment, employ more complicated world models, and preferably use robots able to function even in the absence of direct sensory inputs. The latter would probably involve endowing the robots with artificial emotions, such as fear or curiosity, as discussed in [4]. Further work along these lines is being carried out at present.

REFERENCES

1. M. Wahde and M.G. Nordahl, *Coevolving Pursuit-Evasion Strategies in Open and Confined Regions*, to appear in C. Adami, R. Belew, H. Kitano, and C. Taylor (eds.), Proc. of "Artificial Life VI", Los Angeles, MIT Press, 1998
2. D. Cliff and G.F. Miller, *Co-Evolution of Pursuit and Evasion II: Simulation Methods and Results*, in From Animals to Animats 4, P. Maes et al (eds.), MIT Press, 1996
3. K. Sims, *Evolving 3D Morphology and Behavior by Competition*, in Proceedings of Artificial Life IV, R. Brooks and P. Maes (eds.), Cambridge, MIT Press, 1994
4. S. Pinker, *How the mind works*, W.W. Norton & company, New York, 1997

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Neuroprocessor control systems of intelligent mobile robots

Igor Kaliaev[a], Ivan Rubtsov[b] and Vladimir Chelyshev[b]

[a] Scientific Research Institute of Multiprocessor Computing Systems Supercomputers and Neurocomputers Research Centre 2 Chekhov st., Taganrog, 347928, Russia

[b] Mechatronics and Robotics Department, Moscow Bauman State Technical University2 Baumanskaja st. 5, Moscow, 107005, Russia

The problem of collision-free mobile robot motion in complex real environment without human control is discussed in this paper. The method, oriented on realization by means of neuroprocessor control system (NCS), for such type of problems solution in real time is suggested. The principles of NCS function and structure of mobile robot control system on its basis are designed. The practical realisations of intelligent mobile robot control system on the basis of NCS are described.

1. INTRODUCTION

Today a problem of creation of intelligent mobile robots, which can function autonomously in complex real environment, has a great significance in different branches, for example, in military, in atomic engineering for work in red zone, in space researches and etc. In the first place the complexity of such robots creation connects with the problem of on board control systems construction, which can solve in real time huge complex of intelligent tasks. The problem of robot motion planning and control for achievement of target in beforehand unknown environment takes one of the central places among this complex of tasks. As far as robot environment is a priori unknown it is impossible beforehand to determine and load into robot memory its movement control programs in all possible situations. Therefore, the robot control system must automatically form the current motion control for achievement the target state by optimum way in the rate of environment changes, that is in real time. Such type of problems belongs to class of so called position optimum control problems. The position strategy permits to form the object control a posteriori, correcting it on the basis of additional information, received during motion. The problem of position optimum control has the important practical significance especially in the case, when the object moves under beforehand unknown conditions and influences. The main complexity of position optimum control problems consist in the fact that their solution must be executed in real time of object movement. This requirement hampers the solution of such problems by standard computers with consecutive processing of information. It is explained by difference between sequential way of information processing in computer and parallel nature of position optimum control problems, which sense consists in choosing of one optimum solution among a set of possible alternatives. At the same time the human brain easily solves such type of problems. This fact suggests an idea of using neuroprocessor structures for robot collision-free motion control in real time.

2. THE PRINCIPLES OF NEUROPROCESSOR CONTROL SYSTEMS ORGANIZATION

The method of position optimum control problem reduction to the problem of extremal path construction on the graph-model of the object state space is proposed in [1,2]. The essence of the approach is the following. The object states space $\{Y\}$ is covered by the regular graph $G(Q,X)$, whose vertices $q \in Q$ correspond to discrete points of the space $\{Y\}$ and arcs $x(q_j, q_{j+1}) \in X$ connect the vertices, corresponding to the neighbour discrete points Y_j and Y_j of the space $\{Y\}$. The weight $\gamma(q_j, q_{j+1})$, equalled to the increment of optimisation functional on the transition between points Y_j and Y_{j+1}, is assigned to the arc $x(q_j, q_{j+1})$. Besides, the vertices q_0 and q_k, whose coordinates are defined by the current and target object states, are selected. Moreover, it is necessary to define a set of possible vertices Q_A and set of possible arcs X_A, satisfying to the system of limitations, put on object states and controls.

As a result of such construction we receive the discrete model of the object states space, represented in the form of a homogeneous graph $G(Q, X)$. By using this model the problem of object position optimum control is changed by the problem of path construction on the graph $G(Q, X)$ between the vertices q_0 and q_k. This path must pass only through a set of the possible vertices Q_A and a set of the possible arcs X_A and must have the extremum sum weight of the arcs, belonged to it (Fig.1). It is shown [1,2] that initial arc of this path defines the current vector of optimum control for object collision-free movement to target. The fact that the problem of optimum position control may be reduced to the task of the extreme path construction on the graph-model allows to neuroprocessor control system (NCS) with parallel information processing to solve it.

Let we have a neuroprocessor structure topologically similar to the graph-model of the object states space. Each structure neuroprocessor must correspond to vertex of the graph-model, moreover, if two vertices of the graph are connected by the arc then the neuroprocessors, corresponding to them, are connected by the link. Besides, let the time delay of the signal passing between the neuroprocessor input and its output is proportional to the weight of corresponding arc of the graph. If the arc belongs to the set of impossible arcs X / X_A, then the signal passing through this link is blocked. Similarly, all the links, connecting the neuroprocessor with the neighbour neuroprocessors, are blocked too, if corresponding vertex belongs to the set of the impossible vertices Q / Q_A of graph-model $G(Q, X)$.

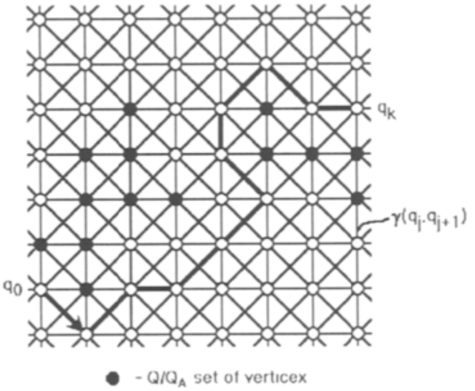


Fig.1. Graph-model of object state space.

If in such structure we shall generate a signal in the neuroprocessor, corresponding to vertex q_k, then this signal will be propagated along the whole structure. In this case the time, during which this signal will reach the inputs of neuroprocessor, corresponding to the vertex q_0, will be proportional to the weight of the extremum path in the graph between the vertices q_0 and q_k. Moreover, the link, by which the first of the signals will come to the neuroprocessor q_0, defines the initial arc of this path.

Analysing the above procedure we may define the following simple functions which each neuroprocessors of such NCS must realise.

1. If the neuroprocessor corresponds to the vertex q_k, then the signal of wave propagation is formed in it.
2. If the signal has come to the neuroprocessor then it is transferred to the neighbour neuroprocessor with the time delay proportionally to weight of the corresponding arc in case when this arc belongs to the set of possible arcs X_A and

corresponding vertex belongs to the set of possible vertices Q_A. Otherwise it is not transferred.

3. In the neuroprocessor q_0 the link is fixed along which the first signal has come to it.

The scheme of NCS neuroprocessor, represented at Fig.2, realises all this functions. Here h_i and $h_i'(i = 1,2,\ldots,l)$ – inputs and outputs of wave propagation signals, α_0 and α_k – inputs of correspondence to vertices q_0 or q_k, α_A – input of correspondence to set of vertices Q/Q_A, $h_i^*(i = 1,2,\ldots,l)$ – outputs for fixation of link index, through which the first signal has come to neuroprocessor, if this neuroprocessor corresponds to vertex q_0.

The proposed NCS has a high vitality, because it ensures the problem solutions when some neuroprocessors or the links between them break down. Really, the break down of some neuroprocessors or links is equivalently to narrowing of allowable conditions area, determinated by system of limitation. In this case the structure ensures the problem solution on remaining part of neuroprocessors and links.

3. ORGANISATIONS OF MOBILE ROBOT MOTION CONTROL SYSTEM ON THE BASES OF NCS

The structure of mobile robot collision-free motion control system, based of NCS, is shown at Fig.3. The information about robot current state and current condition of environment is received with help of navigating subsystems and information subsystem. On the base of this information the graph-model $G(Q,X)$ of robot state space $\{Y\}$ is formed by means of model formation subsystem, i.e. the weight of arcs and sets of possible vertices and possible arcs of graph-model $G(Q,X)$ are determined. Besides with the help of this subsystem the vertices q_0 and q_k, corresponding to current and target robot states are chosen. Note, that the target robot state can be set by operator or systems of top level. The graph-model $G(Q,X)$, formed by means of model formation subsystem, is reflected further in NCS, i.e. the links time delay and signal passing through neuroprocessors are programmed in NCS. If this graph-model does not change during work of system and only the boundary conditions, determined by current and target robot states, are varied, then NCS can be programmed once on the initial stage. If it is not so, then in each cycle of system work the NCS programming must be repeated again.

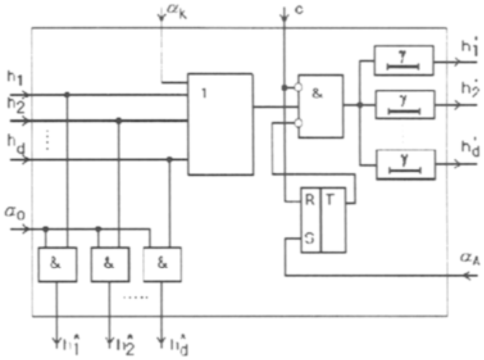


Fig.2. Scheme of NCS neuroprocessor.

Further the wave of signals is formed in the NCS neuroprocessor q_k, corresponding to the target robot state. The inputs of neuroprocessor q_0 are connected to inputs of solution formation subsystem. This subsystem fixes the index of link, through which the first signal reaches neuroprocessor q_0, and then defines the vector of current motion control for achievement the target

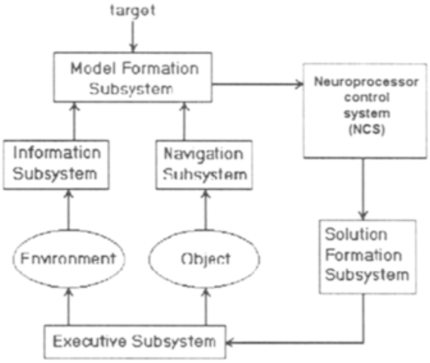

target
Model Formation Subsystem
Neuroprocessor control system (NCS)
Information Subsystem
Navigation Subsystem
Environment
Object
Solution Formation Subsystem
Executive Subsystem

Fig.3. Structure of mobile robot control system.

state by optimum way in current environment conditions. This vector is transmitted further on executive subsystem for realisation.

After this the cycle of system work is repeated anew, with new information about current robot state and current condition of environment and so on, until the current robot state will not coincide with target state.

4. EXPERIMENTAL RESULTS

The theoretical researches have formed the basis for creation of series of experimental intelligent mobile robots (IMR), intended for exploration of Solar system planets surface, in particular Mars. This researches was carried out in frameworks of Russian space program.

The feature of such robots functioning consists in the large time of information exchange between robot and operator (more than 40 minutes if robot works on Mars). Therefore the robot must have the units for automatic motion control for achievement target without operator help. As far as the environment of robot motion is beforehand unknown, then it is impossible beforehand to plan and to keep in robot memory all variants of motion to target. Therefore the problem of position optimum control of robot motion to target in current situation is appeared. In spite of this, the rigid demands are made to the robot control system, such as:

- compactness, giving the possibility of system installation on robot chassis;
- survivability, ensuring preservation of system function in conditions of repair impossibility;
- super high performance, ensuring of robot motion control in real time of situation change (or new information reception about situation).

The control system, based on NCS, answers to all these demands.

With the purpose of compact realization of NCS the special element base, including itself the chip of NCS fragment and multichips module was developed and created. The high homogeneity and simplicity of NCS neuroprocessors permit to put on one chip the large number of such neuroprocessors. Created chip of NCS fragment contains 128 neuroprocessors. Moreover the opportunity of such chips connection one with another with the purpose of NCS size increment is provided. Multichips module contains 8 chips of NCS fragment in self (1024 neuroprocessors at whole) and the opportunity of such modules connection one with another is also provided.

The scheme of IMR planet-rover control system is shown at Fig.4. It contains two processor modules (PM1 and PM2), serving for processing of sensor information from range-finder and TV-camera; processor module (PM3) for processing navigating information; processor module (PM4), serving for IMR movement planning and processor module (PM5), forming control effect on IMR chassis drives for realization of constructed movement with account of chassis dynamic properties. The coordination of all processor module work is executed with help of central processor module.

The functions of accumulation and correction of robot knowledge base about environments are entrusted on central processor too.

The system feature consists in the fact that the processor module PM4, used for task solving of IMR movement planning to target, is constructed on the base of NCS.

The NCS, which is used in processor module PM4, contains 4096 elementary neuroprocessors, incorporated in common solution field and realised by means of 32 chips of NCS fragment. With help of NCS the all possible variants of IMR movement trajectory to target are analysed in parallel and the optimum one is chosen on the base of current information about environment, accumulated to current time moment in robot knowledge base.

As a whole the system works as follows. On the basis of information, which is received in current moment of time by means of sensor (scanning laser range-finder or TV-camera), corresponding processor module builds the model of visible environment zone. The integrated sign, determined difficulty its passing for robot, is put in correspondence for each section of this model. After this sensor processor module forms the request in central processor, which read the constructed model and with help of special algorithm put it over accumulated previously the hierarchical environment model, connected with current IMR position and made up its knowledge base about

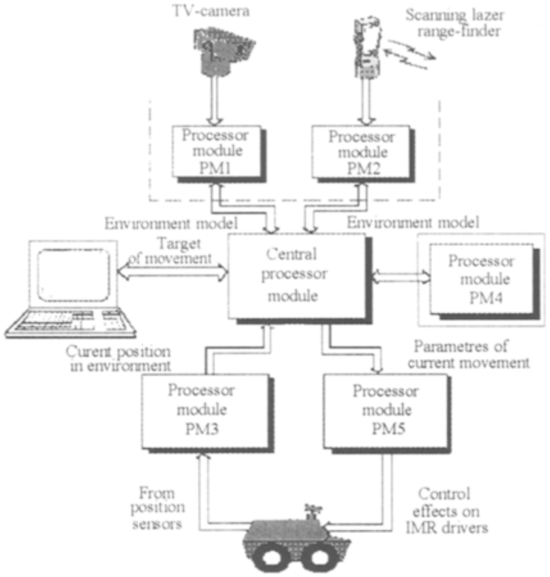

TV-camera
Scanning lazer range-finder
Processor module PM1
Processor module PM2
Environment model
Environment model
Target of movement
Central processor module
Processor module PM4
Curent position in environment
Parametres of current movement
Processor module PM3
Processor module PM5
From position sensors
Control effects on IMR drivers

Fig.4. Structure of IMR control system.

environment. After this central processor passes to main processing procedure, which consists in following. The model of environment, accumulated in knowledge base, is reflected in NCS, where the all possible variants of IMR movement trajectory to target are analysed and is chosen the optimum one, information about which is transmitted back in central processor module. On the basis of this information central processor calculates the parameters of current IMR movement alone the chosen trajectory with account of chassis dynamic characteristics. These parameters are transmitted further to processor module PM5, which is used for control effects of IMR chassis drives for realisation of current movement parameters. The procedure of optimal trajectory planning of movement and controlling action formation on IMR chassis drivers is repeated periodically through each 0,2 sec. The collision free IMR movement through previously unknown real environment with the speed up to 10 km / hour is provided by means of use every time on the base of new information, accumulated to this moment in knowledge base.

The experiments have shown the efficiency of intelligent mobile robot control systems construction on the bases of NCS. Firstly, they are compact, that permits to put them on robot chassis. Secondly, they have large level of survivability as far as they may function when some neuroprocessors and links disable. At last, thirdly, they have of super high performance, achieved by use of non numerical method of solution, that provides the possibility of collision-free robot function in complex real environment without human control.

5. CONCLUSION

The suggested approach can be used for creation of mobile robotic systems, intended for function in beforehand unknown, real environment without human control. For example, it can function in cosmic or underwater, in zone of radioactive and chemical pollution, for inspection of dangerously explosive zones and objects, for realisation of rescue works and so on, as well as industrial and domestic intelligent robots.

REFERENCE

[1]. I.A. Kaliaev. Homogeneous Neurolike Structures for Optimisation Variation Problem Solving.-PARLE'93. Parallel Architectures and Languages Europe. 5-th Inter. Conf., Munich, Germany, June, 1993, pp.438-451.

[2]. I.A. Kaliaev. Homogeneous Neurolike Structures for Solving the Optimum Control Problems.- PARCELLA'94. Proc. of the VI Inter. Workshop by Cellural Automata and Arrays. Potsdam, Germany, Sept. 21-23, 1994, pp.187-195.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Trajectory Control and Time-Varying Stabilization of Nonholonomic Wheeled Mobile Robot: A Hybrid Strategy

A.C. Victorino[a] and P.R.G. Kurka[b] and E.G.O. Nobrega[c]

[a,b]Departamento de Projeto Mecânico, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, Caixa Postal: 6122, CEP: 13083-970, Campinas, SP- Brazil.

[c]Departamento de Mecânica Computacional, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, Caixa Postal: 6122, CEP: 13083-970, Campinas, SP- Brazil.

Nonholonomic Wheeled Mobile Robot (NWMR) is a mechanical system which is characterized by non integrable kinematic constraints. Brockett's rank condition shows that a nonholonomic system cannot be stabilized to a rest configuration by means of smooth state feedback laws. Its structural properties are analyzed and the kinematic state space models necessary for the understanding of the behavior of the NWMR are presented. Trajectory tracking problems for mobile robots by means of feedback linearization is then considered. Dynamic feedback linearization and time-varying feedback are successively applied to handle the tracking of a moving reference trajectory and stabilization of the robot to a rest configuration in a hybrid strategy. The trajectory control of nonholonomic NWMR using feedback linearization is presented, with simulated and experimental results based on a Khepera, a 55 millimeters of diameter miniature mobile robot. Experimental results show the applicability of the theoretical studies presented.

1. INTRODUCTION

The feedback control laws for nonholonomic mobile robots have given rise to a great amount of literature in the recent years. The nonholonomic constraints arise from the assumption that the wheels of the robots roll without slipping. The control problem is to design feedback laws that can stabilize the robots about an equilibrium point. However nonholonomic robots cannot be stabilized to a rest configuration by a smooth state feedback, even though it is completely controllable [1-3]. Some researchers have proposed smooth time-varying state feedback, i. e. control laws depending not only on the state vector but also explicitly on the time variable, such an approach is not a natural sequel to the traditional solutions, but has the great advantage of preserving control smoothness. These are non stationary feedback controls, as pointed out in references [4-5].

We are concerned, in this article, not with the stabilization problem about an equilibrium point, but with the stable tracking of a reference motion. As long as the reference trajectory does not include rest configurations, the tracking problem does not meet Brockett's obstruction and may be achieved by means of smooth state feedback laws [6]. The technique of linearizing completely the equations of the systems by means of static state feedback is not sufficient to handle this problem, but complete linearization can be obtained as long as the robots velocity is different from zero [7].

Experimental tests are unusual in the literature about feedback control of nonholonomic mobile robots. Our propose in the present article is to verify the experimental validation of the control techniques dynamic feedback and the non stationary feedback on the model of the miniature nonholonomic mobile robot Khepera [8]. The complete task asked the robot is evolving from an initial to a final state, following a previously established trajectory. The tracking of the reference trajectory is achieved via dynamic feedback linearization as pointed out in section 3.1; the stabilization to the rest point is performed using time-varying feedback law in a hybrid way as shown in section 3.2 and 3.3.

The article is organized as follows: in section 2 a kinematic model of the robot under consideration is presented. Trajectory tracking control and the stabilization via state feedback is addressed in section 3 where the control problem is divided in two parts. Firstly the static and dynamic state feedback laws are applied to the mobile robot under consideration, secondly the non stationary feedback laws is discussed and simulated. The complete task for the robot is, then, achieved and simulated.

2. CONFIGURATION KINEMATIC MODEL

2.1. Robot Posture

The position of the robot is completely described by vector ξ of posture coordinates:

$$\xi = (x, y, \theta)^T \qquad (1)$$

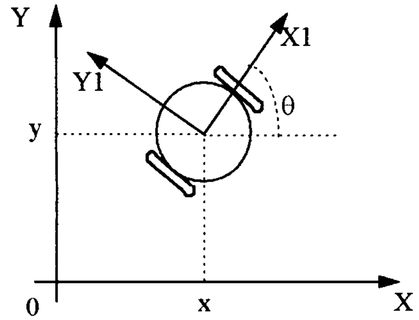


Figure 1: Posture of the mobile robot on the cartesian plan.

where x, y are the coordinates of a point, on the frame attached to an arbitrary basis *{X₁,Y₁}*, with respect to the inercial basis *{X,Y}*, Figure 1.

2.2. Kinematic Model

The mobile robot under consideration has two conventional fixed wheels. It is shown in [9] that the motion of such a robot is described by a kinematic state space model:

$$\dot{x} = B(x)u \qquad (2)$$

where $x \equiv \xi$; $B(x) = R^T(\theta)\Sigma$; $u = \eta$ and $\Sigma = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

A relation between the velocities of the wheels and the input vector η is given by:

$$\eta = D\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix};\; D = \begin{bmatrix} r/2 & r/2 \\ r/2R & -r/2R \end{bmatrix} \qquad (3)$$

where $\dot{q} = \{\dot{q}_1, \dot{q}_2\}$ is a vector of velocities of the wheels, r is the radius of the wheels and 2R is the distance between the wheels. The kinematic model can be rewritten as,

$$\begin{Bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{Bmatrix} = \begin{bmatrix} -\frac{r}{2}sin\theta & -\frac{r}{2}sin\theta \\ \frac{r}{2}\cos\theta & \frac{r}{2}\cos\theta \\ \frac{r}{2R} & -\frac{r}{2R} \end{bmatrix} \begin{Bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{Bmatrix} \qquad (4)$$

3 CONTROL STRATEGIES

The fist part of the control strategy of the robot is a tracking of the reference trajectory and the second part is a stabilization to a final point.

3.1. Trajectory Tracking

Dynamic state feedback laws can handle the tracking of a trajectory which contains no rest points for any mobile robot equipped with two or less steering wheels, as shown in reference [10]. This strategy is applied to the kinematic model of the mobile robot under consideration:

Firstly the static feedback linearization procedure is applied [11-12]:

$$\begin{Bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{Bmatrix} = \begin{bmatrix} -sin\theta & 0 \\ \cos\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} \eta_1 \\ \eta_2 \end{Bmatrix} \qquad (5)$$

$$\varphi = \begin{Bmatrix}\varphi_1\\ \varphi_2\end{Bmatrix} = \begin{Bmatrix} x\\ y\end{Bmatrix} \qquad (6)$$

where φ is the chosen vector of output functions.

Differentiating the output function:

$$\dot{\varphi} = \begin{bmatrix} -sin\theta & 0\\ \cos\theta & 0\end{bmatrix}\begin{Bmatrix}\eta_1\\ 0\end{Bmatrix} = E\begin{Bmatrix}\eta_1\\ 0\end{Bmatrix} \qquad (7)$$

We easily verify the appearance of input η_1, the longitudinal velocity, but matrix E is singular. The dynamic feedback take place and the input vector will then be delayed, generating a new input η_3 defined as: $\dot{\eta}_1 = \eta_3$. This provides the following extended model:

$$\begin{aligned} \dot{x} &= -sin\theta\eta_1\\ \dot{y} &= \cos\theta\eta_1\\ \dot{\theta} &= \eta_2\\ \dot{\eta}_1 &= \eta_3 \end{aligned} \qquad \chi_e = \begin{bmatrix} x\\ y\\ \theta\\ \eta_1\end{bmatrix} \quad u_e = \begin{Bmatrix}\eta_2\\ \eta_3\end{Bmatrix} \qquad (8)$$

where χ_e is the extended state vector, and u_e the new input vector.

Continuing the static feedback linearization, now applied to the extended model yields:

$$\ddot{\varphi} = \begin{bmatrix} -\cos\theta\eta_1 & -sin\theta\\ -sin\theta\eta_1 & \cos\theta\end{bmatrix}\begin{Bmatrix}\eta_2\\ \eta_3\end{Bmatrix},\ \ddot{\varphi} = E(\chi_e)U \qquad (9)$$

It is seen that matrix E is invertible, under the restriction that: $\eta_1 \neq 0$. The longitudinal velocity of the robot must be not null. Choosing a control law U that the cancels the nonlinearities lead to $U = E^{-1}(\chi_e)\nu$, where ν is an auxiliary input vector that will be conveniently designed. Application of such a control law allows the extended model, described in equation (8), to be rewritten as the following linear system,[13]

$$\ddot{\varphi} = \nu \qquad (10)$$

Considering ϕ_{ref} as a reference trajectory to be followed by the output function φ and k_1, k_2 the feedback gains, the auxiliary control ν can be designed as:

$$\nu = \ddot{\varphi}_{ref} - k_2\dot{\tilde{\varphi}} - k_1\tilde{\varphi}; \tilde{\varphi} = \varphi - \varphi_{ref} \qquad (11)$$

where $\tilde{\varphi} = \varphi - \varphi_{ref}$ is the convergence error. Substituting (11) in (10) gives an exponentially convergent dynamic error $\ddot{\tilde{\varphi}} + k_2\dot{\tilde{\varphi}} + k_1\tilde{\varphi} = 0$. Simulation results are presented below.

The reference trajectory to be followed by the robot is:

$$\begin{aligned} x_{ref} &= 0.3 - 0.08t + 0.00565685t^2 +\\ &-0.000377124t^3\\ y_{ref} &= x_{ref} \end{aligned} \qquad (12)$$

such is a non singular trajectory (i. e. $\eta_{ref} \neq 0$, $\forall t \geq 0$);with initial conditions:

$$\begin{aligned} x_{ref}(0) &= y_{ref}(0) = 0.30m\\ \dot{x}_{ref}(0) &= \dot{y}_{ref}(0) = -0.08m/s \end{aligned}$$

The initial condition of the robot is:

$$\begin{aligned} &x(0) = 0.22m; y(0) = 0.30m;\\ &\theta(0) = -3\pi/4rad \end{aligned}$$

This reference trajectory was planned based on the velocity limitations of the motors of the khepera robot, being the maximum velocity permitted for each motor of 127 pulses/10ms, corresponding to 1 m/s. Figures 2 and 3 show the simulation results of the dynamic feedback controller. The convergence errors are asymptotically zero.

3.2. Stabilization to a rest point

To stabilize the robot to a final configuration, in our case the origin of the reference system, a periodic-time control law CL3, proposed in [4], is applied where the input controls are as follows:

$$u_1 = x sin(\theta) - y\cos(\theta)$$
$$u_2 = 40[x sin(t) + u_1 \cos(t) sin(\theta)] - [\theta + 40x\cos(t)]$$

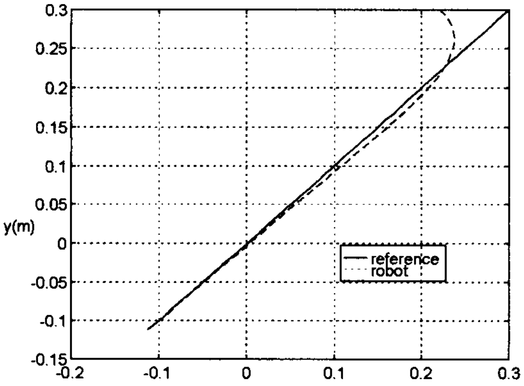


Figure 2. Trajectory tracking.

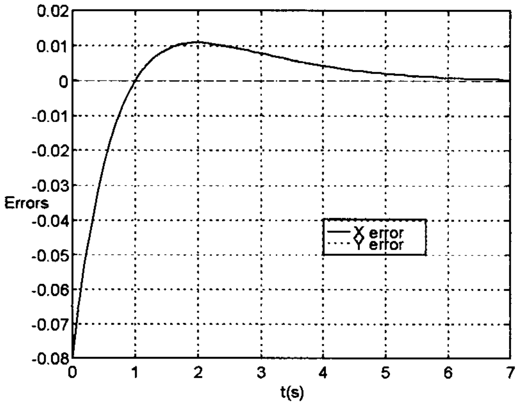


Figure 3. The convergence errors.

The stabilization to the origin is exponential but after some time the velocities are not large enough to move the robot. Figure 4 show the stabilization of the robot with some tolerance.

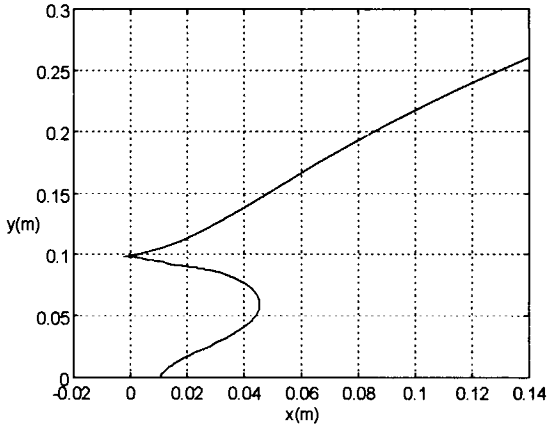


Figure 4. Stabilization of the robot to a rest point.

3.3. A hybrid strategy

Considering reference trajectory with a final stop at the origin, the robot might initially track this trajectory via dynamic feedback. As the robot approaches its goal configurations, a change is made to the time-varying law, in order to bring the robot to the stop point, where dynamic feedback law would tend to infinity, this is a hybrid strategy which combines the dynamic feedback and time-varying laws, taking advantage of each law [7], [10], [14]. The moving reference trajectory that stops at the origin considered here is:

$$x_{ref} = (1/3)\cdot 0.00125\cdot(8-t)^3 \qquad x(0) = 0.14m$$
$$y_{ref} = x_{ref};\dot{x}(0) = \dot{y}(0) \qquad y(0) = 0.26m$$
$$\dot{x}(0) = -0.08m/s$$

Simulation results are shown in Figure 5.

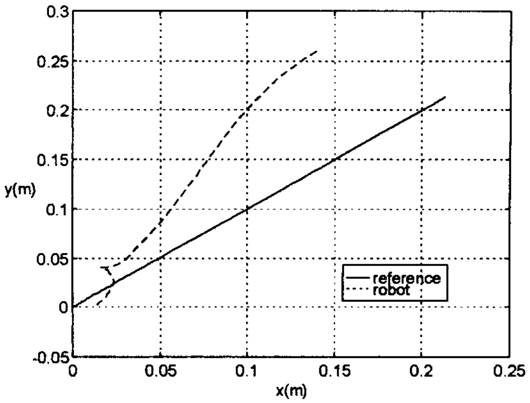


Figure 5. Trajectory tracking and stabilization to a final point: Hybrid strategy.

The commutation to the time-varying law occurs when $x^2 + y^2 \leq 0.032m$, as seen in Figure 5, and the robot can be stabilized to the final point of the trajectory.

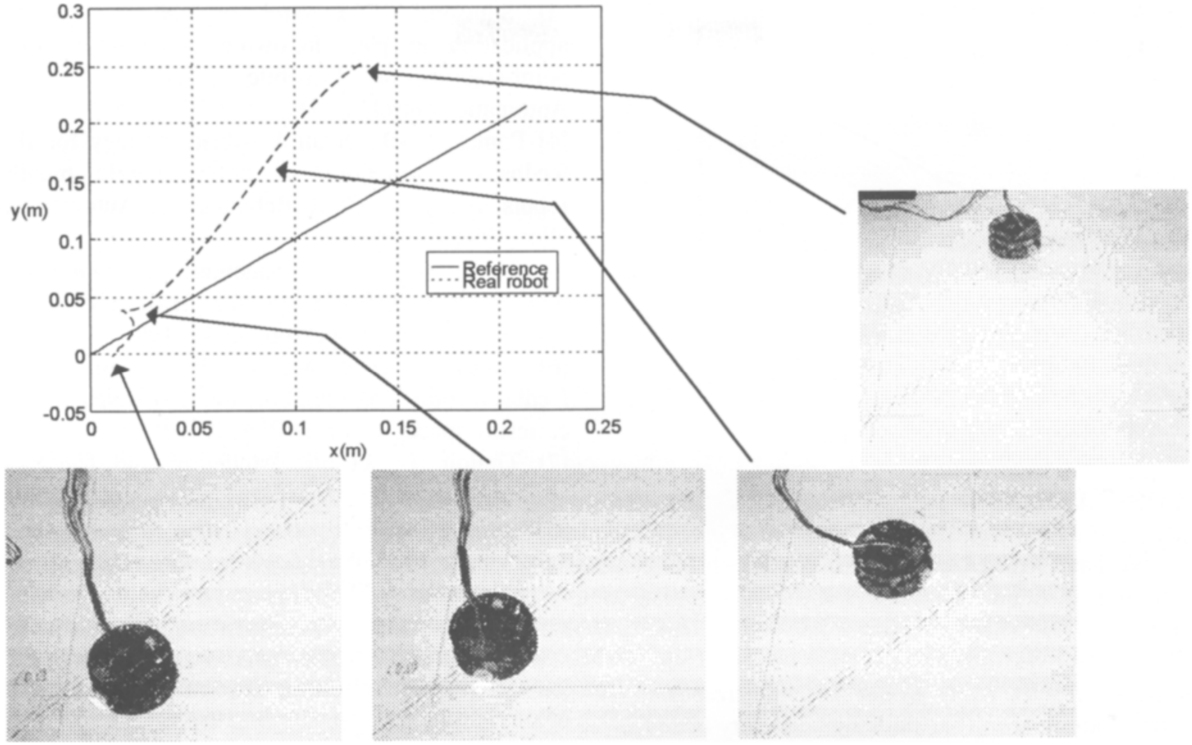


Figure 6. Experimental tracking of a singular trajectory with some positions of the robot.

4. EXPERIMENTAL RESULTS

The experimental work was done to a Khepera [8], a two-wheel miniature robot with 55 mm of diameter; each wheel is driven by an independent DC motor coupled through a 25:1 reduction gear. Its communication with the computer is made using a RS232 serial cable. There are 8 infra-red sensors around it that can indicate the proximity of the obstacles, but those are not used in this work. The control software is a VISUALBASIC program constructed based on the MATLAB programs used in the simulations. Figures 6, 7 and 8 present the experimental results, where it can be noted the proximity with the simulation case.

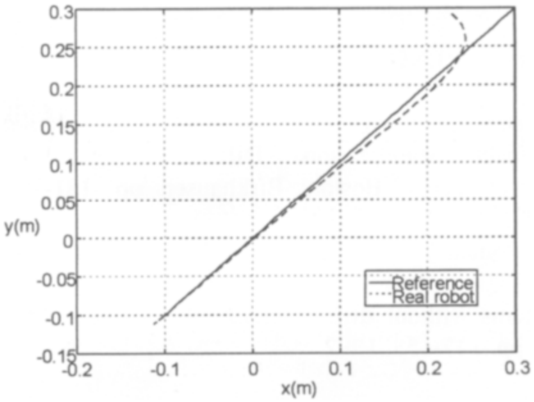


Figure 7. Experimental trajectory tracking

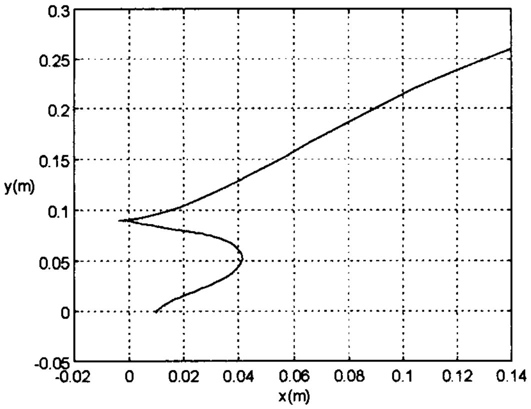


Figure 8. Experimental stabilization.

5. CONCLUSIONS

We have focused on a global control strategy combining dynamic feedback law to track the nonsingular reference trajectory and time-varying feedback law to achieve a final stabilization. Simulation results, using MATLAB, are presented where it can be seen that tracking trajectory errors are exponentially convergent to zero. Experimental tests show the applicability of the control methods presented in this article.

6. REFERENCES

[1] Brockett, R. W.; at all, Asymptotic stability and feedback stabilization, Differential geometric control theory, Boston: Birkhauser. pp. 181-191, 1983.

[2] Pomet, J. -B., Explicit design of time-varying stabilizing control laws for a class of controllable systems without drift, Systems & Control Letters, 18, pp. 147-158, 1992.

[3] Samson, C., Control of chained systems application to path following and time-varying point-stabilization of mobile robots, IEEE Trans. Automatic control, 40(1), pp. 64-77, 1995.

[4] Pomet, J. -B., et all,A hybrid strategy for the feedback stabilization of nonholonomic mobile robots, IEEE Int. Conf. Robotics and Automation, Nice, France, 1992.

[5] Sϕrdarlen, O. J., Exponential stabilization of nonholonomic chained systems. IEEE Trans. on automatic control, 40(1), pp. 35-49, 1995.

[6] Samson, C. et al., Mobile robot control part 1: feedback control of a nonholonomic wheeled cart in cartesian space, Report INRIA, ICARE, 1990.

[7] Thuloit, B., et all., Modeling and feedback control of mobile robots equipped with several steering wheels, IEEE Trans. robotics and automation, 12(3), pp. 375-390, 1996.

[8] Kepera USER MANUAL. Version 4.06, 1995.

[9] Campion, G., et all, Structural properties and classification of kinematic and dynamic models of wheeled mobile robots, IEEE Trans. robotics and automation, 12(1), pp. 47-62., 1996.

[10] D'andrea-Novel, B., Control of nonholonomic wheeled mobile robots by state feedback linearization, The int. Journal of robotics research,14(6), pp.543-559, 1995.

[11] Isidori, A. *Non Linear Control Systems.* New York: Springer-Verlag, 2nd Ed., 1989.

[12] Slotine, J. J. & Li, W., Applied nonlinear control, Englewood Cliffs, NF: Prentice Hall, 1991.

[13] Marino, R., On the largest feedback linearizable subsystem, Systems & Control Letters, 6, pp. 345-351, 1986.

[14] Bloch, A. M. & McClamroch, N. H., Contollability and stabilizability properties of a nonholonomic control system, Proc. 29th Conf. on decision and control, December, 1990.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Visual Guidance of Mobile Robots Using a Neural Network

A.J.Baerveldt, A.Björnberg and M.Gisbert

Centre for Computer Systems Architecture, Halmstad University
Box 823 S-30118 Halmstad, Halmstad

In this paper we present a self-learning method for low-level navigation for autonomous mobile robots, based on a neural network. Both corridor following and obstacle avoidance in indoor environments are successfully managed by the same network. Raw gray scale images of size 32 by 23 pixels are processed one by one by a feed forward neural network. The output signals of the network represent the appropriate steering actions of the robot.

1. INTRODUCTION

Mobile robot navigation is a very difficult task and is today, at a general level, an unsolved issue. The reason why robot navigation is not solved is that the environment is both very hard to model and to recognise, especially changing environments. There have been a lot of methods proposed for robot navigation, e.g. dead reckoning, navigation based on active beacons, on artificial landmarks and a natural landmarks, as reported in [1]. Often these methods need knowledge of the environment and require an intensive engineering effort to switch from one environment to the other. Self-learning is required if we want the robot to adapt easily to different environments. Artificial neural networks is a popular and promising class of learning algorithms and might be appropriate for mobile robot navigation.

The goal of the work presented in this paper is to investigate the suitability of neural networks for guiding an autonomous robot through different corridors, while it avoids people and other obstacles. The only sensor we use is a CCD-camera. It is the intention to combine this later with other sensors like sonar range sensors. However, the motivation of this work is to investigate the possibilities and limitations of a visual sensor only combined with neural networks.

We developed a self-learning method for low-level navigation for autonomous mobile robots, based on an artificial neural network. Both corridor following and obstacle avoidance in indoor environments are managed by the same network. Raw grayscale images of size 32 by 23 pixels are processed one at a time by a feed-forward neural network. The output signals from the network directly control the motor control system of the robot. Our lab reported preliminary results in [2]. The work reported here is an extension where more training examples were used and different network architectures were investigated.

Experiments using neural networks have been done for corridor following as reported in [3]. In that work, however, pre-processing is applied in form of an edge detection operation and a hough-transform

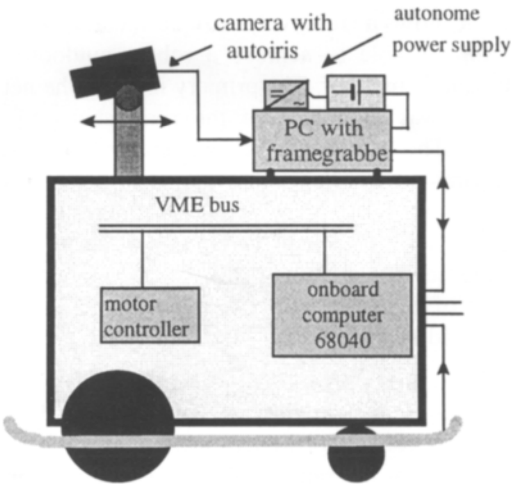


Figure 1 Schematic view of "Robot de Neuro".

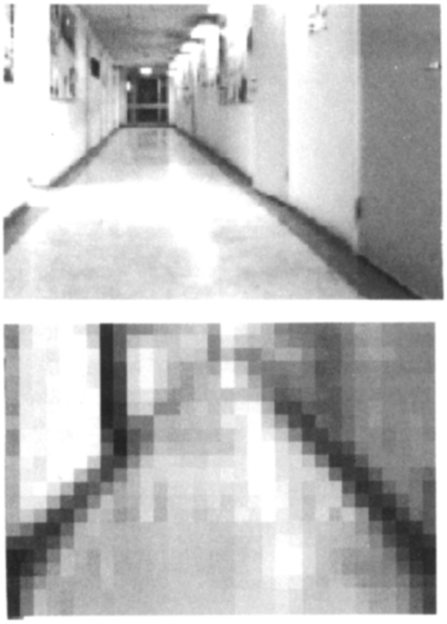


Figure 2 Above the image from the camera and below the sub-sampled image (32x23), which is the input to the neural network.

to detect straight lines. The hough-space is then fed into the neural net. The use of this kind of pre-processing is time-consuming and presumes the presence of a few clear edges in the environment.

Our work is based on the results obtained in the ALVINN project reported in [4] by Pomerleau. He showed how a relatively simple neural network structure is able to achieve a robust outdoor road following behaviour. The primary input to the neural network was an image of the road ahead of the vehicle. The neural network was trained on-line by observing the steering behaviour of a human driver together with the actual image of the road. The training set was complemented by an artificial training set which was created by transforming the images of the road into images not driving straight on the middle of the road.

We do not apply on-line learning but show that a set of images and steering control values is adequate to obtain a robust corridor following behaviour. We also show that the same network architecture can be used to obtain a robust follow-me behaviour, which makes the robot follow people like a dog.

The experimental set-up we used is shown schematically in figure 1. The visual input is provided by a camera equipped with an autoiris lens, which is the only sensor. The mobile robot, called "Robot de Neuro" is a differential drive type.

2. CORRIDOR FOLLOWING BEHAVIOUR

2.1 Neural Network Architecture

The eight bit gray-level image from the camera, sub-sampled to 32 by 23 pixels, was used without any pre-processing as input for the network, as shown in figure 2. The desired steering control values form the output of the network. We arrived at five output nodes to obtain a smooth steering behaviour. Each output node corresponds to one control action, which are strong left, slightly left, straight ahead, slightly right and strong right. The control action is solely determined by the output node with the highest value, i.e. a winner takes all arrangement.

We experimented with the number of nodes in the hidden layer. Five nodes proved to be good for our application. Fewer nodes led to a relatively poor performance. More nodes led to only slight improvements at the cost of slower training. Encouraging results were also obtained without a hidden layer, however only for training data containing clean images of the corridors without any obstacles. For more complex training patterns the network with one hidden layer proved to be better.

The proposed configuration as shown in figure 3 gives a neural network with 32x23 = 736 linear input neurons, five sigmoidal hidden neurons, and five sigmoidal output neurons. This leads to a total of 736x5+5x5 = 3705 weights.

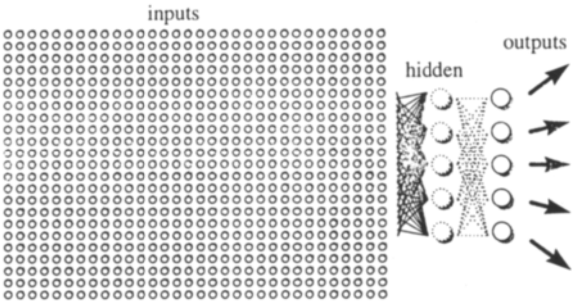
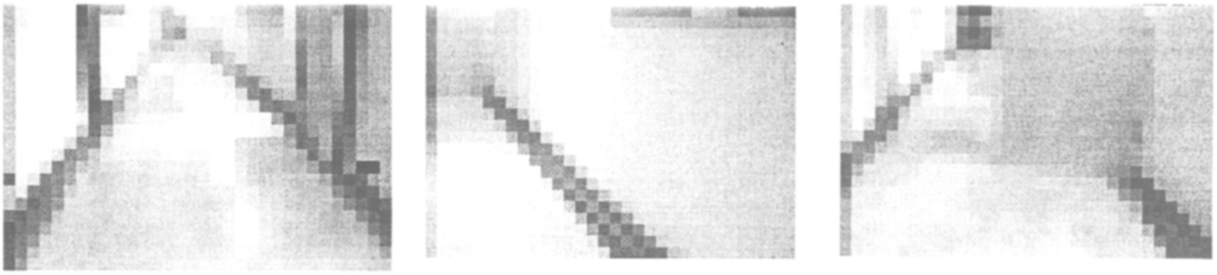


Figure 3 The final architecture of the neural network.



Straight Ahead Slightly to the Left Straight Ahead

Figure 4 Different training examples with the desired control action. On the right an obstacle in the form of an open door can be found.

2.2 Training

One way of speeding up the learning is to use an adaptive learning rate scheme. Therefore, RPROP [5], Resilient Propagation, was chosen. It uses the sign changes in the gradient to adapt the step size. In principle, if the gradient changes sign after an update, we are close to a minima and should reduce the step size. If the gradient has the same sign we should increase the step size (accelerate the learning). A neural network with a large number of free parameters, i.e. weights, and relatively few training examples as we encounter in our case, suffers a big risk of overfitting to the training data. This results in a neural network model not able to generalise well on new data. A common method to avoid "overfitting" is a technique called *early-stopping* [6]. This requires the data set to be divided in two. One set is called the training set and the other the validation set. The network is trained on the training set. Every training epoch the error of the validation set is calculated as well. The training is stopped when the error of the validation set reaches a minimum. Continuing the training would lead to overfitting thus decreasing the error of the training set and increasing the error of the validation set.

To collect training patterns for the neural network, a number of images, showing the different situations the robot should be able cope with, while moving through the corridors, were taken. These situations include situations where the robot:

- is placed at varying distances to the wall
- is placed at different angles
- encounters obstacles, e.g. open doors

The total amount of images that we collected was 500. The images were then classified by a human operator and a desired output is assigned to every image, i.e. one of the five possible steering actions: strong left, slightly left, straight ahead, slightly right, strong right is assigned to each image. In figure 4 some typical training examples are shown. In order to double the available data set and to ensure a symmetric training of the neural network, we flipped all the images in horizontal direction and added these patterns to the existing set. Off course, the corresponding desired steering control action has to be flipped as well. This is demonstrated in figure 5. In this way we obtained a total of 1000 training examples.

We divided this in a training set of 700 and a validation set of 300 patterns to be able to apply the early stopping algorithm to avoid overfitting as described previously. After learning, about 10% of the patterns in the validation set were misclassified. This seems high. However, in practice this was not a problem as one whole cycle from grabbing an image to the output of the network could be done 7 times a

Slightly to the Right *Slightly to the Left*

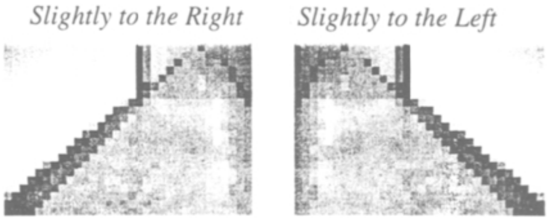


Figure 5 Flipping the training patterns. On the left the original and on the right the flipped one.

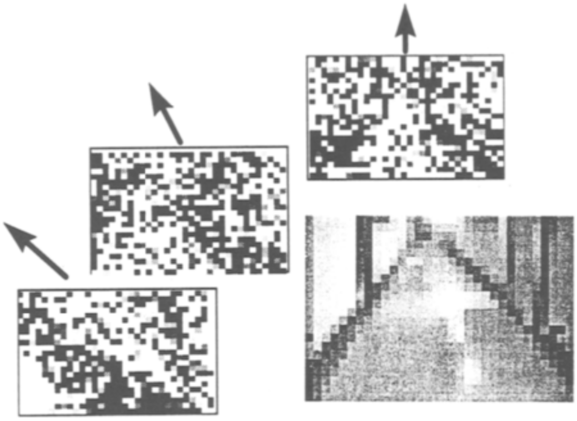


Figure 6 Weight representation from the input layer to the control actions outputs: straight ahead, slightly left, strong left.

second on a standard PC. This resulted in that the robot did not follow perfectly the middle of the corridor due to misclassifications but still was able to follow the corridor robustly while showing a very acceptable behaviour.

2.3 Analysis

To be able to understand how the network processes the image it is instructive to first look at the network *without* a hidden layer. This network has 32 by 23 input neurons and five output neurons and proved to perform well on training data only containing clean images of the corridor without any obstacles. The arrays of weights of this network between the input layer and each output node can be regarded as a representation of the image features which lead to a high activation of the specific output node. Figure 6 shows a visualisation of these weight arrays. Dark areas represent strong sensitivity for dark areas in the image, and white areas strong sensitivity for white areas in the image. The gray pixels represent areas with less sensitivity. In the figure the weights to the output nodes slightly right and strong right are not shown. These are similar to the horizontal flipped weight arrays of the slightly left and strong left nodes respectively. This is due to the fact that the problem is symmetric and due to our training strategy of flipping all training patterns as described in chapter 2.2.

From figure 6 it can be seen how the network reacts. An image in a straight ahead situation is shown and it is easy to see that the weight array for the output node straight ahead matches best. In case of the presence of obstacles it gets more complicated and it is not so clear any more which output node matches best. This is the reason why an additional hidden layer of five nodes gives an improvement for the more complicated cases with obstacles. The layer between the hidden and the output nodes then fuses the responses of the five corridor templates into an appropriate steering action, while the input plus hidden layer acts the same as the network without a hidden layer. This explains also why networks with less then five hidden nodes perform relatively poor as less basic corridor templates can be learned.

2.4 Results

We trained the network and tested the robot with different camera angles, corresponding to a field of view in front of the robot between about 1,5 and 12 meters. With no obstacles in the corridor scenes a robust behaviour was obtained for this wide range of field of view. However, for effective obstacle avoidance the best results were obtained by looking relatively far ahead (ca. 12 meter). During extensive tests, the robot was able to follow the corridor and succeeded in avoiding both static and dynamic obstacles, such as open doors, illustrated in figure 7, and moving humans.

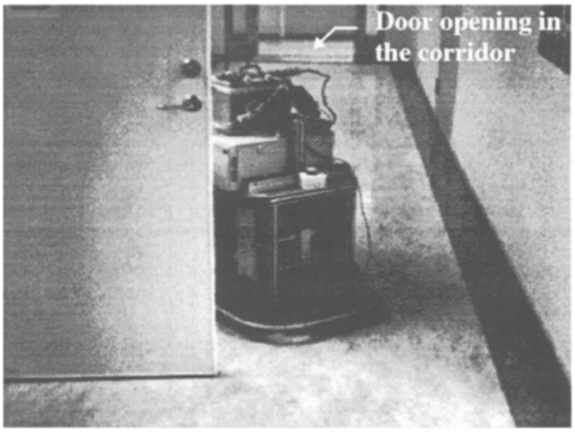


Figure 7 The robot is effectively avoiding the door. At the top: the door opening through which the robot was able to pass collision free.

The robot was first tested in the corridor where it was trained. To illustrate the accuracy, the robot, itself 60 cm width, was able to drive collision free through a door, with a width of 88 cm, situated in the corridor itself. The door opening can be seen in figure 7. Examples of tested, but not trained, environments and situations that the robot managed are: driving in corridors narrower than the one trained in, driving through a corridor with people standing along both walls, following a single wall (i.e. driving through a corridor environment with on one side an open area with tables and chairs). The robot was even able to follow the 90 degrees turn in the corridor, as illustrated in figure 8.

3. FOLLOW ME BEHAVIOUR

The follow me behaviour is intended to make the robot follow a moving person. This behaviour can be invoked when a person meets the robot in the corridor and wants it to follow after to some place. We trained this behaviour with the same network architecture used for the corridor following behaviour. For safety reasons we added an additional output node to regulate the speed. If the person was far away the speed should be increased and if the person was close the speed should be decreased. The performance proved to be good and the robot was able to follow persons both wearing dark as well as light clothes. The robot was able to follow the persons through doors and was even able to make a 180 degree turn in the relatively narrow corridors.

4. CONCLUSIONS

We succeeded in having the robot follow different corridors at Halmstad University, successfully avoiding obstacles and people that tried to stand in the way, and also made it turn to the left or right when the end of a corridor was reached. We also implemented a follow-me behaviour which makes the robot follow people like a dog. This behaviour is learned by the same network architecture as the corridor following behaviour. We may draw the conclusion that neural networks with video images as sensory input constitute a promising self-learning system to achieve robust reactive behaviours for mobile robots.

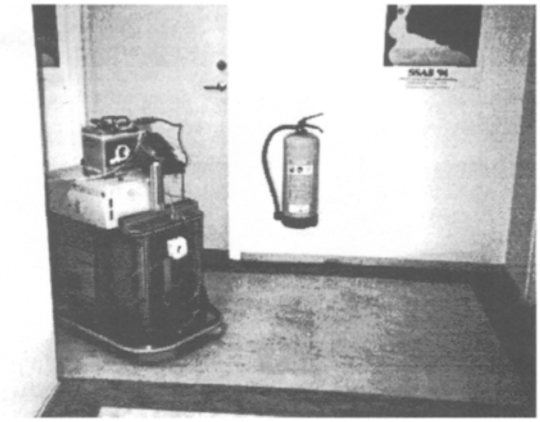


Figure 8 The robot is turning to the right at the end of the corridor into the next corridor.

REFERENCES

1. L. Feng, J. Borenstein and H.R. Everett, "Where am I - Sensors and Methods for Autonomous Mobile Robot Positioning". Technical Report UM-MEAM-9421, University of Michigan, 1996.
2. M. Jonsson, P. Wiberg and N. Wickström, "Vision-Based Low Level Navigation Using a Feed-Forward Neural Network", 2nd Int. Workshop on Mechatronical Computer Systems for Perception and Action, Pisa, Italy, Feb 1997.
3. M. Meng and A. C. Kak, "Mobile robot navigation using neural networks and nonmetrical environment models," IEEE Control Systems, vol. 13, no. 5, pp. 30-39, Oct. 1993.
4. D. A. Pomerleau, "Neural Network Perception for Mobile Robot Guidance". Kluwer Academic Publishers, 1993, ISBN 0-7923-9373-2.
5. M. Riedmiller and H. Braun, "A Direct Method for Faster Backpropagation Learning: The RPROP Algorithm," Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, USA, Mar.28-Apr. 1, 1993.
6. S. Haykin, "Neural Networks - A Comprehensive Foundation", Macmillan College Publishing, 1994.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

CONTROL OF PIEZO ACTUATED MECHANISMS

J. van Dijk

University of Twente, Fac. Mechanical Engineering
p.o. box 217, 7500 AE Enschede, Netherlands
tel: +31-53-4892601 e-mail: j.vandijk@wb.utwente.nl

Abstract

Now a days more and more precision mechanisms are driven by piezo actuators. In these applications the actuators are driven often in an open loop manner. In here the high bandwidth (2000 Hz.) feedback control of a mechanism demanding an accuracy < 20.10^{-9} m is discussed. The mechanism considered is a mechanism for laser bundle manipulation. Two control strategies are compared with respect to there robust performance in the case of model uncertainty. It will be shown that an implementation of a simple PID-control does not provide robust performance and a controller design using the mixed sensitivity approach gives promising results.

1. INTRODUCTION

Due to there high accuracy piezo actuators can be applied with success for driving the precision mechanisms for laser bundle manipulation of $C0_2$ lasers. The piezo actuators are controlled often in an open loop manner. In here the feedback control of a mechanism used for laser welding of surface mounted devices (SMD's) on printed circuit boards (PCB's) is discussed.

The mechanism under consideration has to have a high bandwidth (2000 Hz.) the welding process requires an accuracy better then 20.10^{-9} m at the actuator-tip. The design of this mechanism is discussed in earlier publications [1,2]. In here we discuss two different control strategies of the system.

The control of piezo actuated mechanism is incompatible with the control of mechanisms driven by inductive actuators. The difference is that there is no rigid body mode of the system. There is in this case need for the control of a (number of) vibration modes. The number of vibration modes to control is dependent on the type of control (non co-located co-located), the desired bandwidth and the resonant frequencies of the mechanical subsystem.

An implementation of a digital double lead filter and co-located control using a strain gauge sensor, which is integrated with the actuator, is tested on a set-up. The tests have shown (see Figure 5) that, seen from an accuracy point of view, there is a need for non co-located control.

We show that a non co-located PID-controller combined with voltage driving of the actuator indicates the need for control of higher frequent modes. It will be shown that PID like control is not robust.

The controller design problem is then reformulated as a non- co-located mixed sensitivity control problem and solved using H_∞-optimisation combined with current steering. With the designed H_∞-controller the bandwidth and accuracy is obtained and the system has robust stability.

In the non co-located case an optical sensor at the mirror is applied.

In section 2 a brief description of the designed mechanism is given. Section 3 is devoted to the constitutive equations of the piezo actuator (CMA) and mechanical subsystem models. Section 4 gives a short overview of uncertainty and robustness. Section 5 and 6 contain the actual discussion about the control and the sensors. Finally, section 7 contains the conclusions.

2. DESCRIPTION OF THE MECHANISM

The purpose of the mechanism is to rotate a mirror. The mirror manipulates a pulsating $C0_2$ laser bundle in such away that the bundle can weld the leads of SMD's on a PCB.

The specifications are as follows. In order to weld 200 leads per sec. the time for positioning $t_m = 2\ 10^{-3}$ sec. The positioning accuracy of the welding spot, after t_m, must be ±10 μm. For testing purposes a chip with 20 leads is used. The pitch of the leads is 0.5 mm. Distance between object and mirror is 500 mm. The relative error $U_0 = 10 \cdot 10^{-6} / 5 \cdot 10^{-4} = 0.02$

A bandwidth (lowest eigenfrequency) for the system can be defined, with this bandwidth the system should be able to perform the task within the assumptions and specifications given. The approximated bandwidth can be obtained from [3]:

$$U_0 = \frac{\left(\frac{1}{t_m \cdot f_{BW}}\right)^3}{\pi} \quad (1)$$

In here f_{BW} is the bandwidth of the system in the case the setpoint function is of third order with a triangle acceleration profile and the system's dynamic behaviour is dominant second order. This is a worst-case approximation of the bandwidth for a dominant second order system since it assumes no or very less damping. From (1) the bandwidth specification results in f_{BW}= 1258 Hz. This specification will be used as a the target bandwidth for controller design.

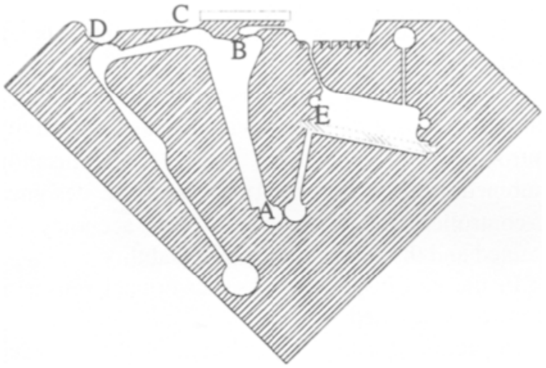


Figure 1: *Design of the mechanism*

The design of the mechanism is shown in Figure 1. The kinematic transfer between mirror rotation (θ) and piezo displacement (x) can be formulated as:

$$i_{mech} = \frac{\theta}{x} = \frac{AB}{BP \cdot AE} \quad (2)$$

In here BP is the distance from joint B to the momentary pole P the mirror is rotating about. The mechanism is designed in such away that the centre of mass of the mirror and support is located in the pole P.

Link	dimension in m
AB	0.05
AE	0.025
BP	0.005
CP	0.015
CD	0.025

Table 1: *Dimensions of the mechanism*

Relation (2) is only valid for small rotation angles of the elements involved. It assumes that the displacement of the pole P is negligible. Kinematic analysis has shown that this relation is not linear, which results in positioning errors in the case of co-located control.

2.1 Sensors

In the test set-up a strain gauge sensor is used co-located with the actuator. Measurements have shown that due to the non-linear kinematics no accurate displacement at the object is obtained. The latter was reason for applying non co-located control with optic sensoring via the mirror. This optic sensor consists of a class 2 Helium Neon laser source and a PSD (position sensitive photodiode) type ShiTek 1L20.

2.2 Modal analysis and measurements

We have applied a modal analysis to the system using ANSYS. The results are shown in Table 2. Measurements (vibration analysis) show great correspondence with the modes and frequencies obtained via de finite element modal analysis.

Mode	FEM	Vibration analysis
1	1469 Hz.	1400 Hz.
2	2960 Hz.	3016 Hz.
3	3999 Hz.	not measured
4	6395 Hz.	not measured.

Table 2: *Results of modal analysis*

The third mode is the first transverse vibration of the actuator. The mirror does not react in this mode. This means that the mode is unobservable and uncontrollable. Consequently, the mode will not have effects on stability of the closed loop system. Its frequency is high enough so it is not expected that this mode will have negative contributions on the accuracy. The modes of vibration are validated by measurements using a Bruel & Kjear dual channel analyser equiped with small piezo acceleration sensors. The results are also shown in Table 2.

3. MODELLING FOR CONTROL

3.1 Constitutive relations of piezo ceramic actuator

The constitutive equations of the applied piezo actuator are:

$$\begin{pmatrix} S \\ D \end{pmatrix} = \begin{pmatrix} s_{11}^E & d_{31} \\ d_{31} & \varepsilon_{33}^T \end{pmatrix} \cdot \begin{pmatrix} T \\ E \end{pmatrix} \quad (3)$$

In here:

S = mechanical strain in m/m
T = mechanical stress in N/m^2
E = electrical field in V/m
D = electrical displacement in C/m^2

s^E_{11} = compliance under condition of a constant electrical field in Pa^{-1} (m^2/N)

d_{31} = Piezo electrical displacement constant in m/V = C/N

ε^T_{33} = permittivity under condition of a constant stress in F/m

Eq. (3) can be expressed in the usual quantities Force F, Charge Q displacement x and voltage U in the following way:

$$(4)\qquad \begin{pmatrix} x \\ Q \end{pmatrix} = \begin{pmatrix} \frac{s_{11} \cdot l}{A_m} & \frac{-d_{31} \cdot l}{\delta} \\ \frac{-d_{31} \cdot A_e}{A_m} & \frac{\varepsilon_{33} \cdot A_e}{\delta} \end{pmatrix} \cdot \begin{pmatrix} F \\ U \end{pmatrix}$$

We have used $E = -\frac{U}{\delta}, D = -\frac{Q}{A_e}$

The new parameters are defined as follows:

l = the length of the piezo actuator

A_e = number of ceramic sheets (n) × b × l with b the width of the actuator

$A_m = b \times h$, h is the height of the piezo actuator

δ = thickness of the ceramic sheet

In the case we wish to have the actuator operating under charge or current control, equation (4) is in the wrong input output form. The form we wish to have in that case is, x and Q as input and U and F as output. This implies an inversion of (4), which results in:

$$(5)\begin{pmatrix} F \\ U \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon_{33} \cdot A_m}{l\{\varepsilon_{33}s_{11} - d_{31}^2\}} & \frac{d_{31} \cdot A_m}{A_e\{\varepsilon_{33}s_{11} - d_{31}^2\}} \\ \frac{d_{31} \cdot \delta}{l\{\varepsilon_{33}s_{11} - d_{31}^2\}} & \frac{s_{11} \cdot \delta}{A_e\{\varepsilon_{33}s_{11} - d_{31}^2\}} \end{pmatrix} \cdot \begin{pmatrix} x \\ Q \end{pmatrix}$$

The insulation between the ceramic sheets does not contribute to A_m, because this material cannot withstand a compressive are tensile force, then $A_m = n \times \delta \times b$ and as a consequence $A_m/A_e = \delta/l$. $\frac{\varepsilon_{33} \cdot A_m}{l\{\varepsilon_{33}s_{11} - d_{31}^2\}} = \frac{A_m}{l \cdot s^D_{11}}$ is commonly expressed as the stiffness k. $\frac{s_{11} \cdot \delta}{A_e\{\varepsilon_{33}s_{11} - d_{31}^2\}} = \frac{\delta}{\varepsilon^S A_e}$ and this is the reciprocal value of the capacity C, with ε^S the permittivity under the condition of a constant strain.

In the case we wish to have the actuator operating under voltage control, x and U are input and Q and F are output. This implies a partial inversion of (4), which results in:

$$(6)\begin{pmatrix} F \\ Q \end{pmatrix} = \begin{pmatrix} \frac{A_m}{l \cdot s_{11}} & \frac{d_{31} \cdot A_m}{s_{11} \cdot \delta} \\ \frac{-d_{31} \cdot A_m}{s_{11} \cdot \delta} & \frac{A_e\{\varepsilon_{33}s_{11} - d_{31}^2\}}{s_{11} \cdot \delta} \end{pmatrix} \cdot \begin{pmatrix} x \\ U \end{pmatrix}$$

Note that the force F in the above equations is the force applied by the load to the actuator. Hence, the force F^* applied to the load is $F^* = -F$!

The parameters of the ceramic multilayer actuator CMA d_{31} are:

Parameters	value	Dim
Y(Young's modulus)	$7 \cdot 10^{10}$	Pa
s_{11}	$1/Y = 1.429 \cdot 10^{-11}$	m^2/N
ε_{33}	$3500 \cdot 8.85 \cdot 10^{-12}$	F/m
d_{31}	$240 \cdot 10^{-12}$	C/N, m/V
n	35	
l	0.038	m
b	0.008	m
d	$40 \cdot 10^{-6}$	m

Table 3 *Parameters of the CMA*

3.2 Modelling the mechanism

The models we are going to use for controller design are a 1DOF and 2 DOF transfer function model. In the 1 DOF mechanical model like in Figure 2, the spring and mass are the equivalent spring stiffness and equivalent mass respectively. F is the force supplied by the actuator. x is the position of the mass.

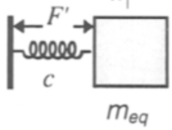


Figure 2

The transfer function model can be written as:

$$(7)\qquad H_1(s) = \frac{x}{F'} = \frac{\frac{1}{c} \cdot \omega_1^2}{s^2 + 2\zeta\omega_1 + \omega_1^2}$$

In here ζ is the relative damping which is 0.008 for low frequencies and 0.004 for high frequencies. ω_1 is de frequency of the first mode of vibration and c is a constant representing the DC-gain. c^{-1} equals F'/x_s in the static equilibrium case with a deformation x_s. In the case x_s is the position of the actuator tip (E), c is approximately equal to the stiffness of the CMA. In the case x_s represents the mirror angle θ, c^{-1} is equal to the inverse of the stiffness of the CMA multiplied by the kinematic transfer i defined in (2).

$$H_2(s) = \frac{x}{F'}$$

$$\text{(8)} \qquad = \frac{\frac{1}{k}\omega_1^2\omega_2^2}{\left(s^2 + 2\zeta\omega_1 + \omega_1^2\right)\left(s^2 + 2\zeta\omega_2 + \omega_2^2\right)}$$

Hence, c is defined with some small parameter uncertainty. Generalisation of (7) results in the transfer function model (8) of a 2 DOF system.

The force F' is delivered by the CMA and can be obtained from (4) in the case of voltage driving

$$\text{(9)} \qquad F' = -\frac{d_{31}\cdot A_m}{ss_{11}\cdot\delta}\cdot k_u \cdot U = d_U \cdot k_U \cdot U$$

and from (5) in the case of current driving.

$$\text{(10)} \qquad F' = -\frac{\dfrac{d_{31}\cdot A_m}{A_e\left\{\varepsilon_{33}s_{11} - d_{31}^2\right\}}\cdot}{s}k_i \cdot i = \frac{d_Q \cdot k_i \cdot i}{s}$$

k_U and k_i in (9) and (10) are the gains of the power amplifier.

4. ROBUSTNESS AND UNCERTAINTY

Although the system has complicated dynamics in high frequencies, these will initially be ignored. Therefore, the controller should have robust performance. This includes that the system should at least satisfy robust stability (RS) [6 Ch 7]. RS can be analysed when we have an upper bound on the uncertainty. If the uncertainty is modelled as unstructured multiplicative uncertainty, the following robust stability condition must be satisfied:

$$\text{(11)} \qquad |\Delta_m| < \frac{1}{|T|}$$

In here T is the closed loop transfer function or complimentary sensitivity function. Δ_m is the uncertainty transfer function. Eq. (11) is obtained using the small gain theorem [4, 6 Ch 7]. The multiplicative uncertainty transfer function can be written as:

$$\text{(12)} \qquad \Delta_m(s) = \frac{K(s)\underline{G}(s) - K(s)G(s)}{K(s)G(s)}$$

Δ_m is derived from:

$$\Delta_m(s) = \frac{\underline{G}(s)}{G(s)} - 1$$

In here $\underline{G}(s)$ is the perturbed open loop transfer function and $G(s)$ is the nominal open loop transfer function. Δ_m should give an upper bound on the uncertainty. Hence, if the nominal model is a 1 DOF model the perturbed model is a 2 DOF model. Because the second mode is then the worst-case (destabilising) uncertainty. Consequently, if the nominal model is a 2 DOF model the perturbed model is a 3 DOF model.

Robust performance for SiSo systems is obtained when the nominal model performs the required specifications, the robust stability criterion is satisfied and M_S is small [5 ch 2]. M_S is the maximum of the sensitivity function $S(s)=1/(1+K(s)G(s))$. The inverse of $S(j\omega)$ is a measure of how far the nominal Nyquist plot $L(j\omega)=KG$ is away from the point (-1,0). Hence, $1/M_S$ indicates the minimal distance from $L(j\omega)$ to the point (-1,0), implicating a safeguard against poor performance due to uncertainty. Also $|S|$ should be small in the frequency regions of the disturbances $G_d{\cdot}d$, because the control error $e = S(G_d d - r)$, with G_d the transfer from disturbance d to the output. r is the reference, so $S(0)$ should be zero to have a zero static control error e.

5. DESIGN OF PID-CONTROLLED SYSTEM

5.1 Voltage driving

In this case we use as nominal model a 1 DOF model of the mechanism. The transferfunction from input (U) to output (x), which is the position of the actuator tip is:

$$\text{(13)} \qquad G_U(s) = \frac{x}{U} = -\frac{\frac{d_U \cdot k_a}{k}\omega_1^2}{(s^2 + \omega_1^2)}$$

The transferfunction of the PID-controller applied is:

$$\text{(14)} \qquad K(s) = k_p \cdot \frac{s\tau_z + 1}{s\tau_p + 1}\cdot\left(1 + \frac{1}{\tau_i s}\right)$$

The I-action is necessary to have $S(0)=0$. In order to have a fast enough response the choice is that the zeros are that far away from the origin that a dominant real pole of the system is assigned between the zeros and the origin. The design is then as follows: The dominant closed loop real pole should be at approximately 500 Hz. (3000 rad/sec).

The zeros are in 6000 rad/sec, the real pole of the tame D-action in 60000 rad/sec and a proportional gain $k_p = 2^1$ gives closed loop poles: −26235 + 25882j, −2.6235 − 25882j, −24529, −3000. Figure 3 shows the response of the closed loop nominal model $\frac{K_{PID}G_u}{1+K_{PID}G_u}$ to a third order setpoint function. Shown is the displacement of the actuator tip. The system behaves reasonably well.

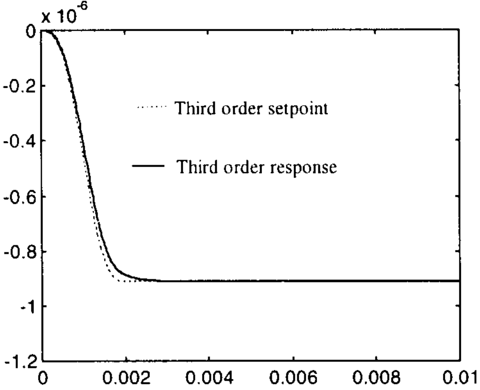


Figure 3: *Response of 1 DOF model*

Figure 4 shows the inverse of the complimentary sensitivity (*T*) and the transferfunction Δ_m. Δ_m is derived using a 2 DOF model Clearly condition (11) is violated.

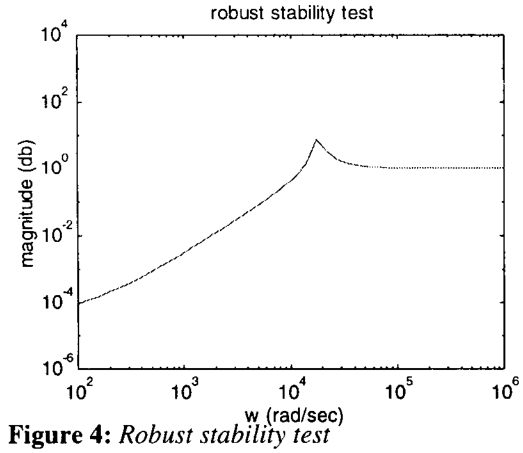


Figure 4: *Robust stability test*

The bandwidth is high, approximately 40000 rad/sec (6000 Hz.). This indicates that some of the higher modes are within the bandwidth. The closed loop poles of te 2 DOF model with this controller are then: −78616, 11528 + 21824i, 11528 − 21824i, −10727 + 9868i, −10727 − 9868i, −2985. We can conclude that PID-control will not work sufficiently in the non co-located control situation.

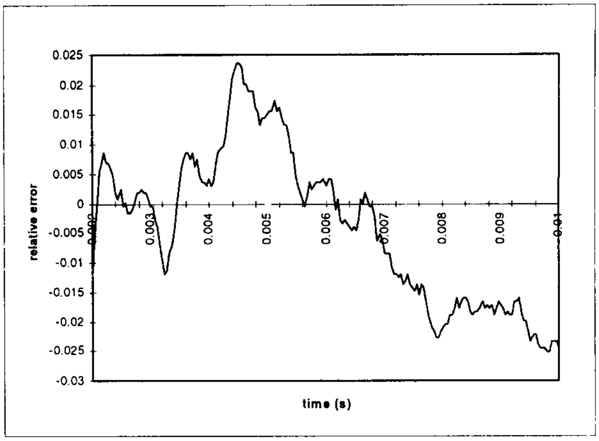


Figure 5: *Relative error after t_m= 2msec.*

5.2 Current driving

The equivalent of PID-control in the case of current driving is the application of a double lead-filter as controller. The design will not be discussed here in detail due to limited space. The accuracy obtained is, due to the non linear kinematics and amplification of process and measurement noise, hardly accurate. which is shown in Figure 5. The accuracy limit is exceeded during welding time. These results required a redesign of the controller. Issues that should be considered in the redesign are the control of the one higher mode and non co-location of sensor and actuator.

6. NON CO-LOCATED CONTROLLER DESIGN

The degrees of freedom for the design of a controller which also controls the third mode, the second mode is non controllable mode and does not have influence on the stability, requires an optimisation procedure. Reformulating the problem as a mixed sensitivity problem opens the possibility to determine the controller using the H_∞ paradigm.

Kwakernaak [6-7] has shown that minimising the norm (15) is a special case of the standard H_∞ problem.

$$\textbf{(15)} \qquad \left\| \begin{bmatrix} W_1 SV \\ W_2 UV \end{bmatrix} \right\|_\infty$$

In here *S* is the nominal sensitivity function as defined in section 4. *U* is the nominal input sensitivity function :

---

[1] The voltage amplifier gain Ka = 4.8.10^6 the dimension is V/μm

$$U = \frac{K}{I + KG} \tag{16}$$

with process transfer G and controller transfer K. U is the transfer function from the disturbance to the process input. U is related to the complementary sensitivity T as $T = G.U$.

Figure 6 shows the configuration and the sense of the weighting filters W_1 and W_2 and the shaping filter V. The mixed sensitivity problem may be used for simultaneous shaping the sensitivity S and the input sensitivity U.

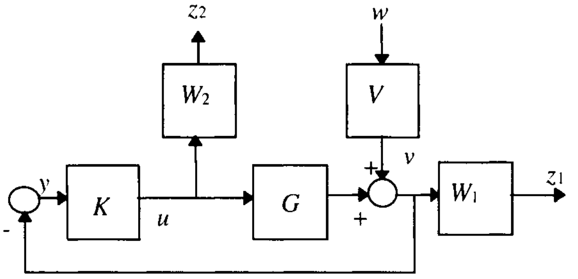


Figure 6: *The mixed sensitivity problem*

In [7] it was shown that, in case of a SiSo system, minimising the criterion (15) is equal to minimising the peak value, which is constant, of the frequency dependent function:

$$\sup_{\omega \in R}\left(\left|W_1 S V\right|^2 + \left|W_2 S V\right|^2\right) = \gamma^2 \tag{17}$$

If the weighting functions are suitably chosen such that $W_1 \cdot V$ is large at low frequencies and $W_2 \cdot V$ is large at high frequencies, then the solution of the mixed sensitivity problem has the property that the first term of the criterion (17) dominates at low frequencies and the second at high frequencies resulting in:

$$\begin{aligned} S(j\omega) &\approx \frac{\gamma}{\left|W_1(j\omega)V(j\omega)\right|}, \text{ for } \omega \text{ small} \\ U(j\omega) &\approx \frac{\gamma}{\left|W_2(j\omega)V(j\omega)\right|}, \text{ for } \omega \text{ large} \end{aligned} \tag{18}$$

This result allows quite effective control over the shape of the sensitivity and input sensitivity functions and, hence, over the performance of the feedback system. In [7] it is also shown that by letting the shaping function V be $V = M/D$, pole placement of the dominant closed loop poles is achieved. This partial pole placement requires that D is the polynomial of the denominator of the process transfer G and M is a polynomial of the same degree as D representing the desired closed loop dominant poles.

6.1 Selection of dominant closed loop poles

We have chosen for the closed loop pole locations the poles of the 5th order prototype transfer function that minimises the ITAE criterion. In order to move the open loop poles not too far we choose that the bandwidth is m times 10000 rad/sec. The ITEA transfer function poles are then: m[-8955,-3764+12920j,-3764-12920j,-5758+5339j,-5758-5339j]. The factor m shifts the bandwidth. Figure 7 shows the influences of m on the complementary sensitivity T, the best results are obtained with m = 1.25

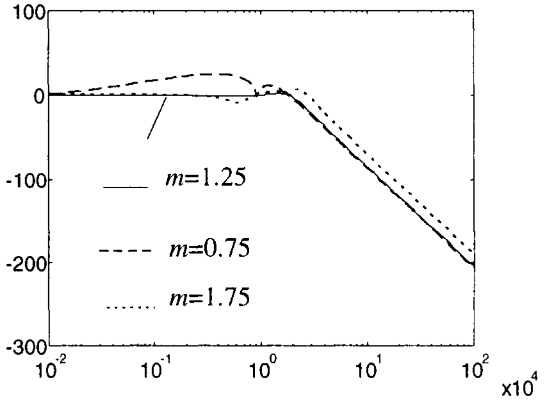


Figure 7: *Influence of m on T*

6.2 Choice of the weighting functions

Because the order of both polynomials M and D is equal, V is asymptotically constant at low frequencies and 1 at high frequencies. Consequently,

$$U(j\omega) \approx \frac{\gamma}{\left|W_2(j\omega)\right|}, \text{ for } \omega \text{ large} \tag{19}$$

Consider the choice of W_2 as:

$$W_2 = c(1 + rs) \tag{20}$$

In here r and c are both constants. If r=0 the high frequency roll-off of U=0 and the roll-off of the complimentary sensitivity T is equal to the roll-off of G. Robustness against high-frequency perturbations (like higher resonance modes) can be improved by making T decreases faster at high-frequencies. This can be accomplished by choosing r≠0. Then U rolls-off at –20 db/decade and T rolls off with the roll-off degree of G –20 db/decade.

The usual form of the weight W_1 is:

$$W_1 = \frac{\omega_1 + s}{s} \tag{21}$$

With W_1 the low frequent behaviour of the sensitivity function S can be shaped. We wish to have rejection of low frequent disturbances at the process input. The process here is the mechanism with transfer function H_2 as in (8). The transfer $R(s)$ from disturbance d to output x in Figure 8 is then:

$$R(s) = \frac{x(s)}{d(s)} = S(s) \cdot H_2(s) \tag{22}$$

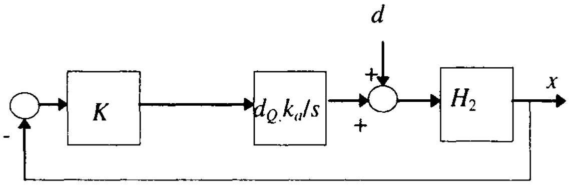


Figure 8

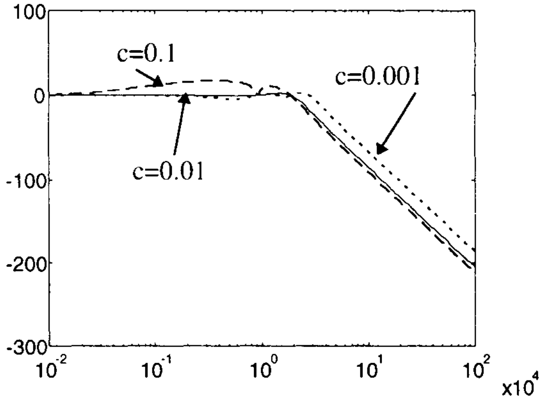


Figure 9: *Influence of r on T*

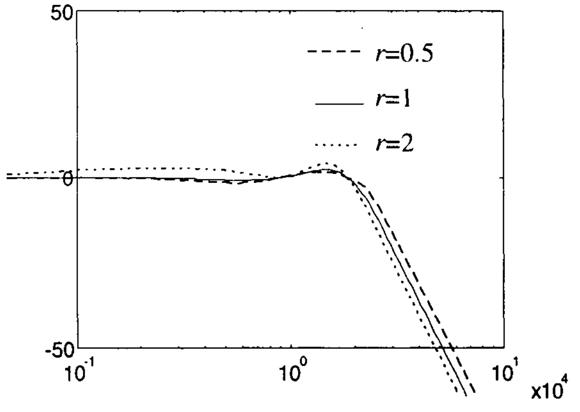


Figure 10: *Influences of c on T*

For low frequencies H_2 is constant and as a consequence R behaves as S. For proper disturbance rejection R and thus S should behave as s in the low frequency range. Eq. (18) shows that this means that the product of the weighting functions $W_1 V$ should behave as $1/s$. $V=M/D$, and D already contains the factor s, due to the integration from current to charge, and M (closed loop polynomial) does certainly not contain the factor s. Hence, we choose $W_1 = 1$.

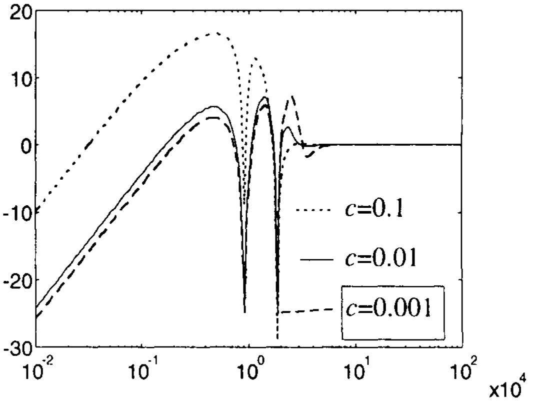


Figure 11: *Influence of c on S*

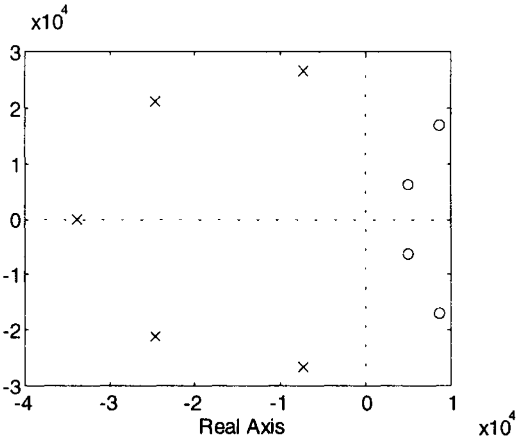


Figure 12: *Pole zero map of controller*

Figure 9 and Figure 10 show the influences of r and c of the weighting W_2 on T. By choosing r bigger the roll-off sets in at lower frequency. Choosing r=10000 rad/sec. gives the best results.

Figure 10 and Figure 11 show that as c increases, T decreases but S increases. A good choice is c=0.01.

The poles and zero's of the designed controller, with c=0.01, r= 1 and m=1.25, are shown Figure 12. The controller is a double notch with an additional pole. The zero's of the notches are in the right half plane.

Figure 13 shows the time response of the model with the controller as indicated and the proces model H_2, due to the right half plane zero's we have shortened the setup time of the input function from 2 msec to 1 msec.

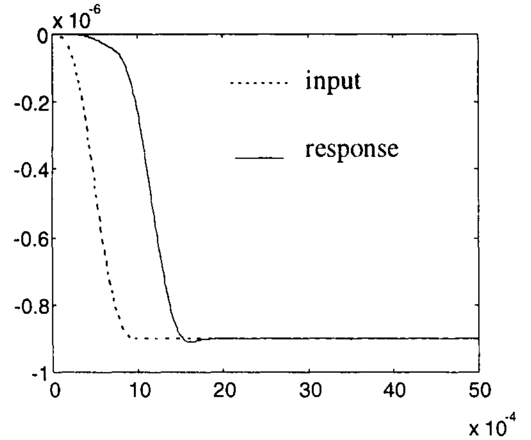


Figure 13: *Reference and response, 2 Dof model*

The result of application of the robust stability condition (11) is shown in Figure 14. Δ_m is derived using the 3 DOF model:

(23)

$$H_3(s) = \frac{x}{F'}$$

$$= \frac{\frac{1}{k}\omega_1^2\omega_2^2\omega_4^2\left(s^2+\zeta\omega_m+\omega_m^2\right)}{\omega_m^2\left(s^2+2\zeta\omega_1+\omega_1^2\right)\left(s^2+2\zeta\omega_2+\omega_2^2\right)\left(s^2+2\zeta\omega_4+\omega_4^2\right)}$$

In here

$$\frac{\left(s^2+2\zeta\omega_m+\omega_m^2\right)}{\left(s^2+2\zeta\omega_m+\omega_4^2\right)}\cdot\frac{\omega_4^2}{\omega_m^2}$$

is the transfer function of the fourth mode, with ω_m representing the zero's and ω_4 representing the frequency of the fourth mode. The zero's of the transfer function are found using a modal analyses with clamped mirror surface.

The value for the relative damping ζ=0.004 is obtained from the vibration measurements. We can conclude that the controller designed using the mixed sensitivity approach looks promising. The controllers were computed using a MATLAB package for the solution of H_∞ optimisation problems [8].

7. CONCLUSIONS

For the feedback control of a piezo actuated tilting mirror mechanism, PID-like control is implemented. In the case of co-located control this control strategy is less accurate. The lag of accuracy is due to kinematic non linearity's and to a too high bandwidth. Shown is that in the case of non co-located control PID-like control is not robust. The problem of designing a controller for the non co-located case is reformulated as a mixed sensitivity problem and solved using the H_∞ paradigm. The designed controller is robustly stable and is promising with respect to the accuracy requirements.

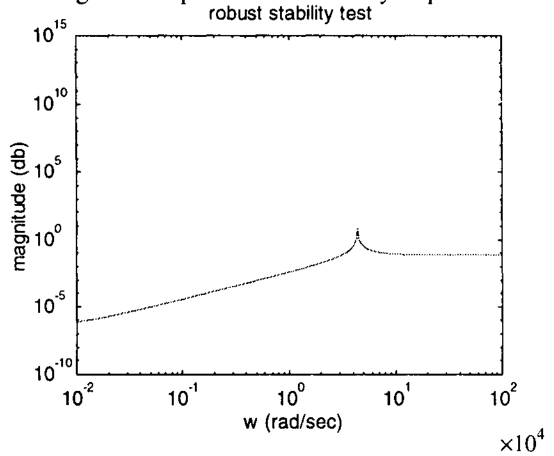


Figure 14: *Robust stability test*

REFERENCES

[1] van Dijk, J., "Mechatronic design of piezo driven tilting mirror mechanism", Jubileumcongres NVPT Leusden, 2 pp. 1997.

[2] van der Berg, R.,"Tilting mirror",Msc-thesis Twente University, n0 WA521, Enschede the Netherlands, 1997.

[3] Koster M.P.,W. van Luenen, T. de Vries,."Mechatronica", lecture notes, University of Twente, Enschede, Netherlands, 1998.

[4] Franklin, G.F. J.D. Powell, A. Enami-Naeini. "Feedback control of dynamic systems", Addison Wesley, New York 1994

[5] Skogestad, S., I. Postlethwaite, "Multvariable feedback control", Wiley New York, 1996.

[6] Kwakernaak, H, "Minimax frequency domain performance and robustness optimization of linear feedback systems", IEEE trans. on auto. control 30, 994-1004, 1985.

[7] Kwakernaak, H, "The polynomial approach to H_∞ optimal regulation", In E. Mosca and L. Pandolfi (eds), Lecture notes 1496. Springer, Heidelberg, 1990.

[8]. Kwakernaak, H., "Matlab macros for polynomial H_∞ -control system optimization. Memorandum n0 881, Dep. Appl. Math. University of Twente, Enschede, Netherlands,1990.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Resonantly Driven Piezoelectric Micropump

(Fabrication of a micropump with the size less than 1cm^3)

Jung-Ho PARK[a], Kazuhiro YOSHIDA[b] and Shinichi YOKOTA[b]

[a]Graduate School, Tokyo Institute of Technology,
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, JAPAN

[b]Precision and Intelligence Laboratory, Tokyo Institute of Technology,
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, JAPAN

As a power source for practical micromachines using fluid power such as in-pipe mobile micromachines, micropumps having high power density are required. In this paper, a piezoelectric micropump using resonance drive is fabricated with the size of $\phi\, 9 \times 10$mm. It basically consists of a bellows as a flexible pump chamber, a piezoelectric actuator for oscillating the bellows, and cantilever type of two check valves. Frequency characteristics and load characteristics of pressure-dependent flow rate are experimentally investigated with the values of various additional mass and valve thickness. As a result, the feasibility of the piezoelectric micropump using resonance drive is confirmed.

1. INTRODUCTION

With recent progress in micromachining technologies, new functional micromachines such as micro robots for maintenance of small diameter pipes of nuclear reactors have been demanded and investigated [1-2]. Microactuators with relatively high power are required for higher performances of such micromachines. Through a theoretical analysis on miniaturization characteristics of various conventional actuators, it was shown by a part of the authors that positive-displacement type actuator using fluid power is excellent in point of power density at a range from 10μm to 10mm [3]. Micro-sized power sources as well as microactuators are required to realize such micromachines. Especially, as a power source for practical micromachines using fluid power such as in-pipe mobile micromachines [4], micropumps having high power density are required for supply of sufficient powers to the microactuators.

It is significant to employ powerful and miniaturizable actuators for realization of such micropumps. Through comparison with other actuators, it is ascertained that a multilayered piezoelectric actuator has high generation pressure and high response. However, its displacement is too small to achieve enough performances. In this study, therefore, magnification of a displacement of the piezoelectric actuator is tried by using resonance drive of a flexible pump chamber.

In the previous paper [5], the piezoelectric micropump using resonance drive was proposed and a large model was fabricated. Through basic experiments on frequency characteristics and load characteristics, a feasibility of the piezoelectric micropump using resonance drive was confirmed.

In this paper, a micropump with the size of $\phi\, 9 \times 10$mm for practical application is fabricated and basic characteristics on pump performances are experimentally investigated, which becomes a criterion for evaluation on optimal structure.

2. THE PIEZOELECTRIC MICROPUMP

2.1. Working principle

Figure 1 shows the working principle of the proposed piezoelectric micropump using resonance drive. An electro-deposited bellows is used as a flexible pump chamber, which expands and contracts in accordance with inner pressure change. As the actuator for oscillating the bellows, a multilayered piezoelectric actuator is employed.

A working principle of the micropump is as follows: The piezoelectric actuator is driven with resonant frequency of the system. Therefore, the resonantly excited oscillating bellows induces large periodic volume and pressure changes in the chamber. With expansion of the bellows, the fluid flows through the inlet into the chamber. On the other hand, when the bellows contracts, the fluid flows through the outlet out of the chamber. Through this process, it is expected that an enlargement of output power and an improvement of pump efficiency can be realized with resonance drive. In this study, to obtain further performances, an additional mass is attached to the free end of the piezoelectric actuator. As the mechanism is very simple and compact, it has a potential for further miniaturization.

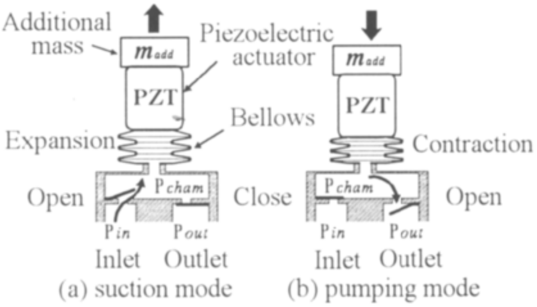


Figure 1. Working principle

2.2. Fabrication of the piezoelectric micropump

Figure 2 shows the schematic of the fabricated micropump. It basically consists of a bellows as a reciprocating chamber, a piezoelectric actuator for oscillating the bellows and normally closed cantilever type of two check valves.

A material of piezoelectric actuator is PZT(Pb[Zr-Ti]O_3). PZT is a standard material of the piezoelectric ceramics. The commercially available PZT actuator used in this study has the size of 3.5 × 4.5 × 5mm. When 100V is applied, it has a maximum displacement of 3 ± 1.5μm with no restraint and a maximum generation force of 130N with no displacement. A mass of the piezoelectric actuator is 0.42g. The bellows is mainly made of nickel and the size is 6.2mm in outer diameter, 3.6mm in inner diameter and 2.0mm in free length. By using epoxy-resin adhesive, the PZT actuator is attached to the bellows. The bellows is expanded and contracted by the PZT actuator and changes the inner pressure. In this system using resonance drive, the resonant frequency of the check valves must be higher than that of the actuation unit. Therefore, the valves are designed to have sufficient stiffness. The natural frequency of about 6kHz with the valve thickness of 20μm is estimated by simple analysis. The valves are made of nickel and fabricated by electroforming technology to obtain sufficient accuracy. The valve size of cantilever type is 1.6 × 0.8 × 0.02mm.

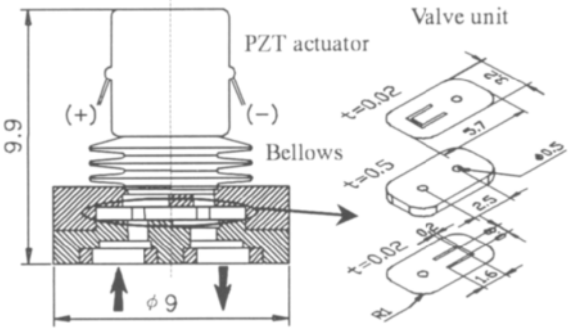


Figure 2. Schematic of the fabricated micropump

3. BASIC EXPERIMENTS

3.1. Experimental apparatus

The experimental apparatus for investigating basic characteristics of the fabricated piezoelectric micropump is shown in Fig.3. In this study, both characteristics with no load pressure and with load pressure are considered for investigating pump performances. Firstly, the measurement of the flow rate with no load pressure is performed by using measuring cylinder and digital balance (resolution : 0.02g). On the other hand, the flow rate with load pressure is measured by using a load bellows or a variable throttle valve. To measure the tip displacement of the additional mass without affecting the state of resonance, non-contact gap detector of capacitance type (measuring range : 0~ 0.2mm, resolution : 0.1μm, response frequency : 2..7kHz) is used. A sampling time for data acquisition is 2ms ~ 2μs. A semiconductor type pressure transducer (measuring range : 0~2MPa) is installed to measure the pressure at the output port. Due to the very high resonant frequency of PZT actuator, a response of PZT actuator strongly depends on dynamic characteristics of the driving amplifier. Therefore, the amplifier for driving the PZT actuator is developed to have sufficient bandwidth of about 10kHz. Tap water is used as a working fluid in this study.

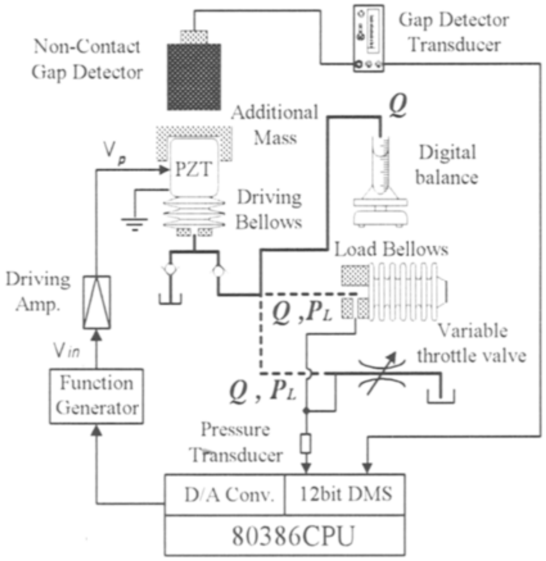

Non-Contact
Gap Detector
Gap Detector
Transducer
Additional
Mass
Q
V_p
PZT
Digital
balance
Driving
Bellows
Load Bellows
Driving
Amp.
Q, P_L
Variable
throttle valve
V_{in}
Function
Generator
Q, P_L
Pressure
Transducer
D/A Conv.
12bit DMS
80386CPU

Figure 3. Experimental apparatus

3.2. Static characteristics

Static characteristics between the applied voltage to the PZT actuator and the tip displacement of additional mass with 38% mass (0.16g) of the PZT actuator are investigated. An example of the experimental results is shown in Fig.4. From the results, it is ascertained that the displacement of about 2.1μm is obtained when the applied voltage V_p changes statically from 10V to 100V.

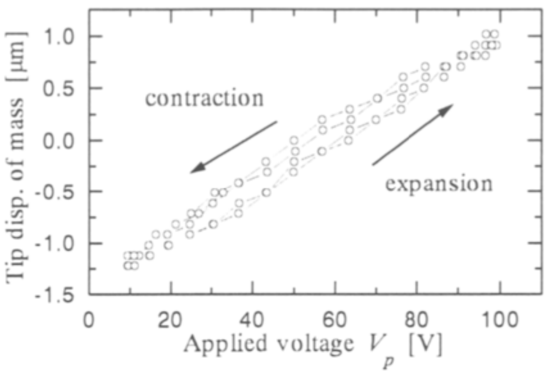


Figure 4. Static characteristics
(voltage vs. displacement)

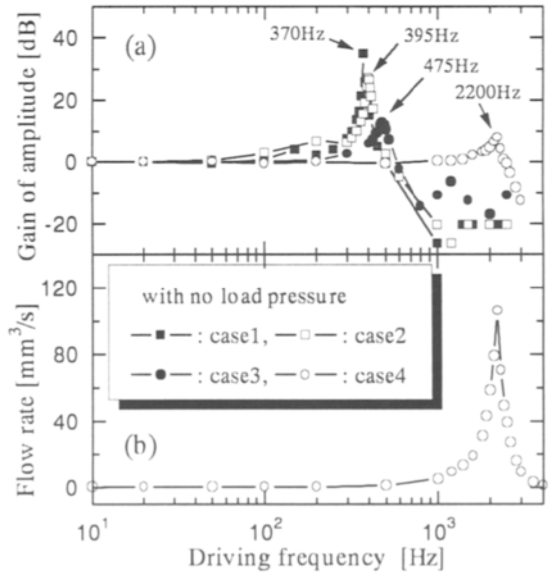


Figure 5. Frequency characteristics when load pressure is zero and additonal mass is 330%(1.38g). (frequency vs. (a) disp. gain and (b) flow rate)

Table 1.
Relation between experimental conditions and resonant frequency when load pressure is zero

	Working fluid	Check valves	Resonant frequency [Hz]	Amplitude of disp. [μm]
Case1	without	without	370	117.3 (34.8dB)
Case2	without	with	395	48.6 (27.1dB)
Case3	with	without	475	9.3 (12.7dB)
Case4	with	with	2200	5.3 (7.9dB)

3.3. Frequency characteristics

Firstly, frequency characteristics under various experimental conditions with no load pressure are experimentally investigated when the amount of additional mass is 330% mass (1.38g) of the PZT actuator. This additional mass is attached to the free end of the piezoelectric actuator having mass of 0.42g. Through pump characteristics with different amount of additional masses in the previous paper [5], it was ascertained that the additional mass of about 300% of the actuator mass is suitable for higher performances. Figures 5 (a) and (b) show the results of the frequency-dependent amplitude gain and the frequency-dependent flow rate under conditions presented in Table 1. In Fig.5 (a), each amplitude gain of the tip displacement of additional mass is obtained as a ratio between the amplitude at each driving frequency and 2.1μm at 10Hz. In Fig.5 (b), the flow rate is measured on an average by the weight of the outflow in 30s.

In the case1 without both working fluid and check valves, the resonant frequency and the amplitude of displacement are 370Hz and about 117μm, respectively. The decrease of the amplitude shown in the case2~4 is thought to be due to the damping effect caused by a fluid and check valves. The increase of resonant frequency is thought to be due to the increase of equivalent spring constant caused by the bellows filled with the working fluid.

From the results in the case4, it is ascertained that the maximum flow rate is obtained at resonant frequency of displacement and this tendency is in good agreement with the results obtained from the large model [5]. As a result, it is ascertained that the maximum flow rate of $106\,\mathrm{mm}^3/\mathrm{s}$ is obtained at 2200Hz with additional 330% mass when load pressure is zero.

4. LOAD CHARACTERISTICS

Next, load characteristics of pressure-dependent flow rate are also experimentally investigated when a bellows or a variable throttle valve is attached to the output port. For this system using water as a working fluid, an influence of bubbles is discharged by boiling the load bellows and an effect of compressibility is neglected. Therefore, when the bellows is employed as a varying load, a simplified equation to measure the flow rate Q is obtained as follows :

$$Q = A\frac{dy}{dt}, \tag{1}$$

where A and y are the effective sectional area and the displacement of the bellows. If the effective sectional area of the bellows is accurately calibrated and the elastic force of the bellows has no external force besides the force PA, the flow rate can be also calculated from the measured pressure P as follows :

$$Q = \frac{A^2}{k}\frac{dP}{dt}. \tag{2}$$

In the previous paper [5], it was experimentally ascertained that the value of flow rate obtained by the displacement is good agreement with that obtained by the pressure. Therefore, the equation (2) is used to measure the flow rate in this paper. Dimensional parameters of load bellows are as follows : the outer diameter is 12.3mm, the free length is 23.6mm, the spring constant is 2.94N/mm and the effective sectional area is $79\,\mathrm{mm}^2$.

4.1. With various additional masses

Firstly, load characteristics with different amount of additional mass are investigated with the load bellows. The valve thickness is 20μm. Figure 6 shows the measured pressures at the resonant frequencies with three additional masses of 180%, 330% and 760%. The driving time is 60s. The applied voltage to the PZT actuator is 55 ± 45V of rectangular wave form. From the results, it is ascertained that firstly the pressure increases gradually with the time and then maintains some pressure, i.e. the maximum pumping pressure at the state.

The flow rates can be calculated with these pressure data by using equation (2). The load characteristics of pressure-dependent flow rate are shown in Fig.7. It is shown that same degree of flow rates are obtained with no load pressure at each additional mass. However, with the increase of load pressure, the result using additional mass of 330% has a better performance than the others. The maximum power with the 330% mass is 1.8mW. Through a comparison with the results of the large model [5], it is ascertained that the adequate mass attached to the piezoelectric actuator is effective for higher performance of this type micropump.

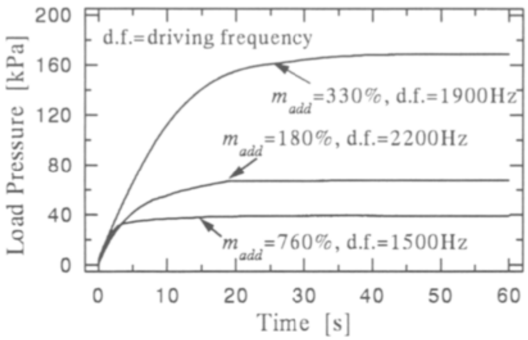


Figure 6. Time history of the pressure measured at each resonant frequency.

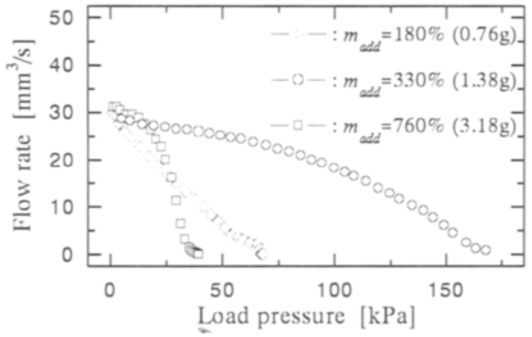


Figure 7. Load characteristics of pressure-dependent flow rate at each resonant frequency.

4.2. With various valve thickness

With the same conditions except for valve thickness, load characteristics with different amount of valve thickness are experimentally investigated. In these experiments, the value of additional mass is 330% (1.38g) mass which is the optimal value derived in the previous section. Figure 8 shows the pressures measured at each resonant frequency with the values of valve thickness of 10μm, 20μm and 30μm. All resonant frequencies with different valve thickness are the same. That is, the change of thickness doesn't influence a resonant frequency of the system although the performances are different.

From the results, it is ascertained that the characteristics of 10μm and 30μm show very unstable characteristics, although its reproducibility is very good. It is in the case of 20μm that the stable characteristic is obtained in all frequency ranges. The results of 10μm and 30μm is thought to be due to the phase shift caused by a stiffness of the valves.

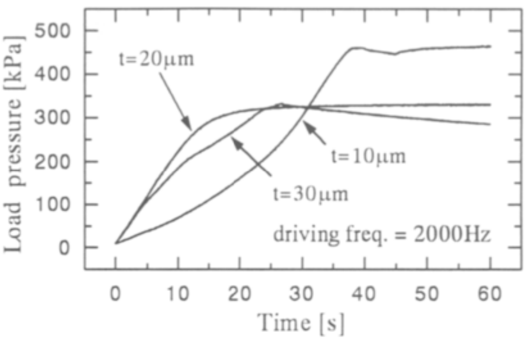


Figure 8. Time history of the pressure measured at each resonant frequency.

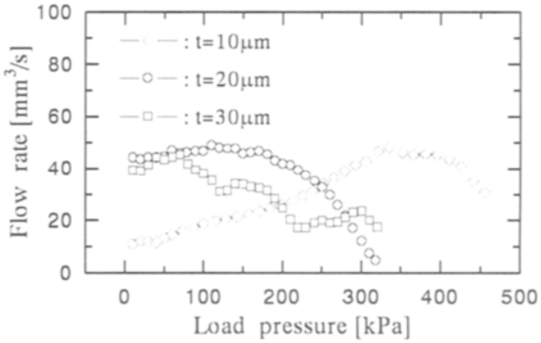


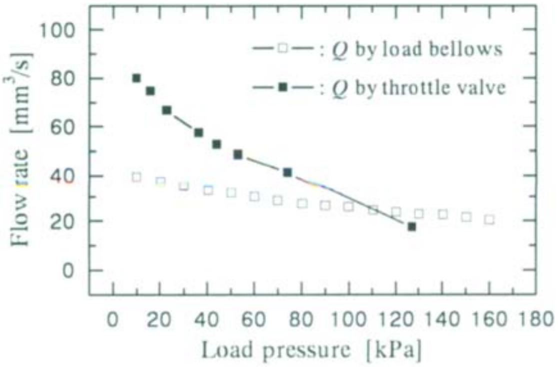
Figure 9. Load characteristics of pressure-dependent flow rate at each resonant frequency.

Load characteristics are shown in Fig.9. From the results, it is ascertained that the maximum output power of about 8.7mW is obtained with the valve thickness of 20μm and adequately chosen valve thickness is effective for stable performance.

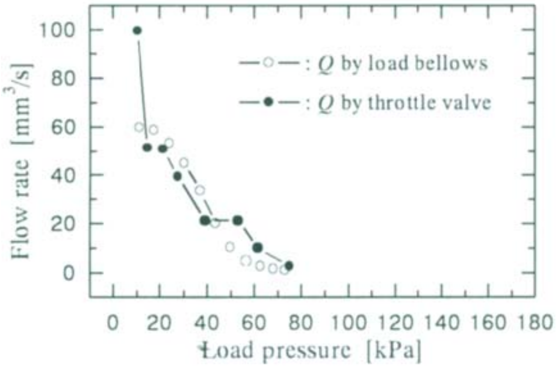
4.3. Comparison of load characteristics

Next, load characteristics using a variable throttle valve as a load are investigated and compared with the characteristics using the bellows. The experiments are performed with the additional mass of 330% and the valve thickness of 20μm. The flow rate is measured on an average by the weight of the outflow in a steady state of the load pressure.

The examples of driving frequencies of 2000Hz and 2200Hz among the experimental results are shown in Figs.10 (a) and (b). The driving frequency of 2200Hz is ascertained as a resonant frequency in the characteristics with no load pressure. On the other hand , the 2000Hz shows better performances



(a) Driving frequency of 2000Hz



(b) Driving frequency of 2200Hz

Figure 10. Comparison results of load characteristics when the additional mass and the values of valve thickness are 330% and 20μm.

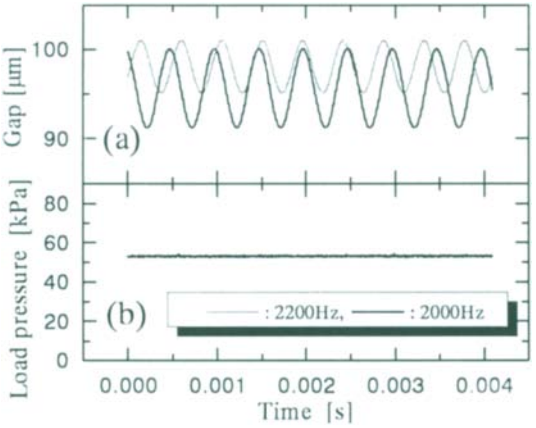


Figure 11. Time history of the measured (a) tip displacements and (b) load pressures.

in the experiments on the load characteristics. In this point, both characteristics obtained by the variable throttle valve and the load bellows show the same tendency. It is also ascertained that the flow rates obtained by the throttle valve are larger than that obtained by the load bellows in low pressure ranges. The reason is thought to be due to the disregarded inertia term of the bellows and so on, but it needs more considerations.

Figure 11 shows the tip displacements and load pressures measured at 2000Hz and 2200Hz. The sampling frequency is 500kHz. In the characteristics without load pressure, it has been ascertained that the amplitude of 2200Hz is larger than that of 2000Hz. However, the amplitude of 2000Hz with some load pressure shows the larger magnitude than that of 2200Hz as shown in Fig.11, and then larger flow rate is obtained although the frequency is low. As a result, pump characteristics having resonant frequency which is changed with the load pressure is obtained, but it has not yet been cleared why the phenomenon occurs.

5. CONCLUSIONS

In this paper, a piezoelectric micropump using resonance drive with the size of $\phi 9 \times 10$mm is fabricated and the characteristics are experimentally investigated. Through basic experiments, it is ascertained that adequately chosen additional mass and valve thickness are effective for higher performance of this type micropump. As a result, maximum flow rate of 80mm^3/s, maximum pumping pressure of 0.32MPa, maximum power of 8.7mW are obtained at the driving frequency of 2000Hz when the additional mass is 330% of the piezoelectric actuator. Owing to its very simple structure, practical application as a power source for micro fluid systems as well as micromachines using fluid power can be expected.

REFERENCES

1. Proc, 10th IEEE MEMS Workshop, 1997.
2. Y. Tatsue, J. the Robotics Society of Japan, 1994, 12-4, pp.508-513 (in Japanese).
3. K. Yoshida and S. Yokota, Proc. FLOMEKO'93, 1993, vol-1, pp.122-130.
4. K. Yoshida, *et al*, Proc. Third JHPS International Symposium on Fluid Power, 1996, pp.229-234.
5. J.-H. Park, *et al*, Proc. ASME IMECE'97, 1997, FPST-Vol.4, pp.77-82.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Piezoelectric final-control elements in mechatronic devices

Dr. A. Boshliakov, Dr. A.Dubin, Dr. I.Rubtsov, Dr. A. Shtcherbin.

This paper report about piezoelectric final-control elements – piezoactuators and piezomotors – developed for make up more wide range of its used. Ways of increase piezoactuators movement range and effective ways of piezomotors control are considered.

1. INTRODUCTION

At present time, simultaneous with the intensive development of microelectronics, some backlog in creation of the final-control elements for mechatronic devices has come to light. In particular, the problem of creation precision and miniature of actuators for linear and limited angular moving is rather topical. Is not less significant topical in precision drives without reduction gear. In the most complete measure to such requirements piezoelectric final-control elements (PFCE) satisfy. In particular, it piezoactuators (PA) and piezomotors (PM), consisting, in a general case, a piezoelectric element (PE) and simple design. PFCE have unique set of positive properties: high resolution, small sizes, large output forces, high efficiency, wide range of working frequencies, absence of an external electromagnetic field, wide temperature range, high reliability and even work in vacuum.

2. LINEAR PIEZOACTUATOR

In particular, on base PA are developed and successfully function such systems as: ultra-precision guide mechanism, cutting-error compensation, lathe actuators for make up piston, oil-pressure servo valve, VTR head, protection vibration active system and other [1], [2], [3].

However, PA has some lacks. The basic lack is a small and limited range of output linear movement. Usual range of linear movement is, on different estimations, 50-100 μm and there is not actuators, which potentially can provide precision, high-dynamics rotation movement up to $\pm 10^0$. It much reduces potential area of use PA. But we know a several task with 250-500 μm wide of linear movement range and up to $\pm 5^0$-$\pm 10^0$ wide of rotation movement range. In the majority of cases, PA has superfluous output force reserve. We can transform this reserve to increase output movement. Improvement of the characteristics will allow increasing PA application area.

2.1. Existing variants of increase PA movement range

For PA, at the moment, the increase of a movement range comes true, in basic:

- By increase the active part length of PA;
- By use intermediate mechanical amplifier;
- By used "step-by-step" principle.

The first variant, as practice shows, is limited following: a) by increase of electrical capacity PA; b) by possible loss of stability PA, as a compressed core; c) by design limits, when required PA has length, exceeding admissible overall dimension.

The second variant results in occurrence additional stiffness loss, backlash, more complicated of a design, increase of overall dimensions, reduction of working frequency range.

The third variant allows beyond all bounds to increase a range of moving, but essentially complexity of a design and control such PA grows. "Step-by-step" PA has small output force and speed.

2.2. Offered way of increase PA movement range

The essential expansion of PA area use can be achieved by increase of a PA output movement range. For this purpose offered to use an inclined arrangement PE concerning a direction of its output movement. In such variant superfluous stock of force will be transformed into increase of output movement. The gear attitude or factor of movement increase such mechanism is approximating equal toa sine of a corner between PA and perpendicular to a direction of output movement. The mechanical characteristic PA has an essentially smaller inclination, becoming more "soft".

2.3. Design variants

On the basis of offered way is developed three constructive variants, each of which has some peculiarities, determining its application area.

The first constructive variant I (Fig. 1) is the simplest. It used one PE 1, inclined located. PE mount in knife-edge support 4 between body 6 and output element 2. Output element move in guide mechanism 5. The preload supplies by elastic element 3. We apply this variant in task for linear movement of load with low frequencies and at the increased requirements to weight and to dimensions of the final-control element.

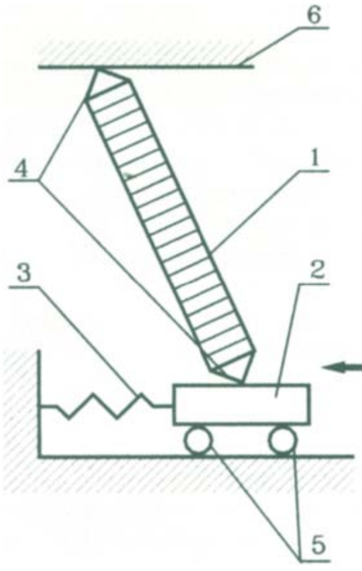


Figure 1. Variant I.

Variant II (Fig. 2) is more combined. It used two PE 1 and 2. They mounted in knife-edge support 4 between body 9 and output element 5 and located V-manner. The preload supplies by adjusting bolts 7 and 8. Such PA has more wide frequencies range in comparison with variant I, because it more stiffness. But this PA has more large dimensions, weight and more complex power amplifier.

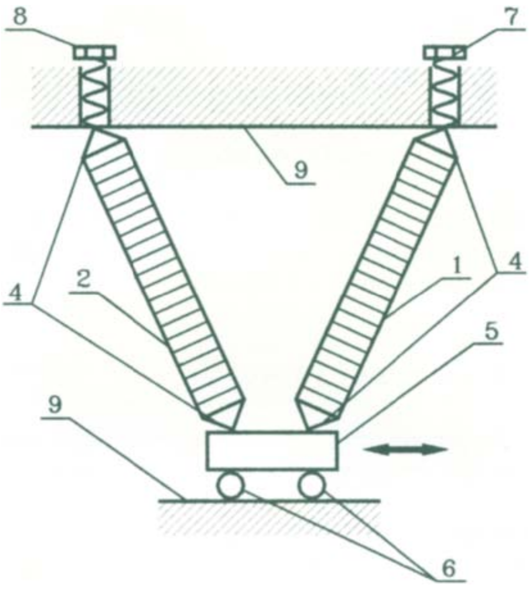


Figure 2. Variant II.

Variant III (Fig. 3) in basic is identical to variant II, but guide mechanism replace by routable support 6. This variant has the more complex design, due to transformation of linear movement to rotation. Its application is expedient in task of a limited load rotation.

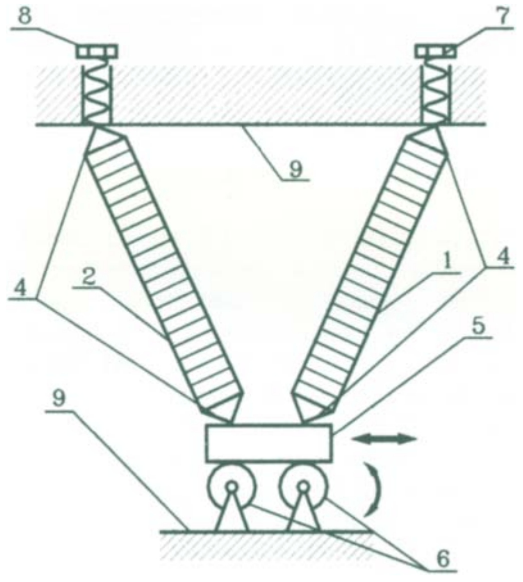


Figure 3. Variant III.

2.4. Dynamic mathematical models and PA samples

The mathematical models (complete, simplificat, linear) of such actuators are developed. Design techniques of piezoactuator are developed. Software-hardware complex, realising the given design technique, is developed. It will allow carrying out as numerical modelling at the stage of designing PA and with high reliability to predict its parameters, and essentially to automate of actuator sample research.

The modelling results, using all three dynamic mathematical models and experimental data's are show at figure 7 as amplitude-frequency characteristic. The curve 2 is corresponding to linear dynamic mathematical model, the curve 3 – simplificat, the curve 4 – complete. The experimental points are mark by symbol «*».

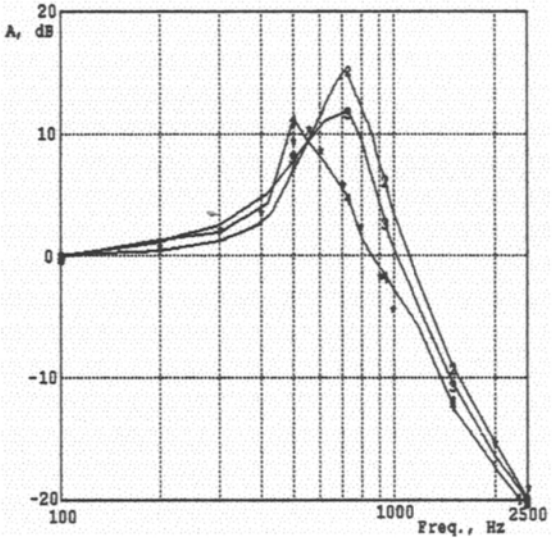


Figure 4. Amplitude-frequency characteristic of piezoactuator.

The PA response to step input is show on figure 7. It is obvious, that the device has oscillatory properties.

On actuators samples were obtaining increased PE output movement 10-12 time, and output rotation $\pm 5^0$. The figure 6 is show photo of PA sample, which based on variant II. The PA used two PE, which have 180 H of output force and 20 µm of output movement. The PA has 15 H of output force and 240 µm of output movement. The figure 7 is show photo of PA sample, which based on variant III. The PA also used two PE, which have 180 H of output force and 20 µm of output movement. The PA output shaft has $\pm 5^0$ of output rotation.

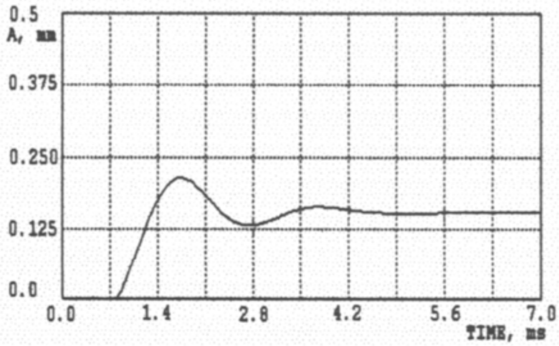


Figure 5. Response to step input.

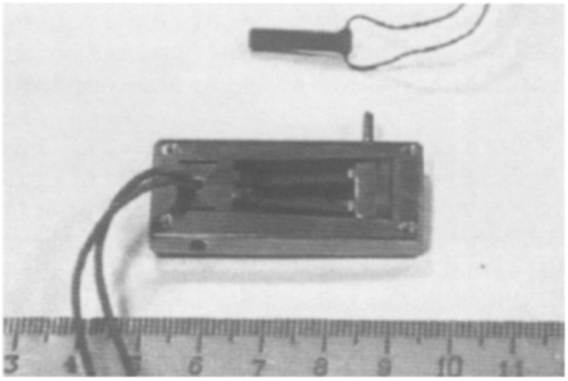


Figure 6. Piezoactuator sample bases on variant II.

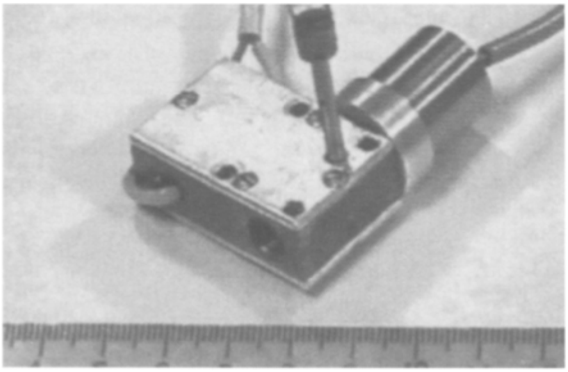


Figure 7. Piezoactuator sample bases on variant III.

3. PIEZOMOTOR

PM distinguishes:

- Design simplicity, easy to produce, defined lack of windings, and, as corollary of it, high reliability;
- Small rotary speeds of the drive output shaft and rather large developed moment on comparison with electromagnetic drives allow to refuse a reduction gearbox for many applications;
- Fast action of the motor allows to create high dynamic drive on base of it (with wide working frequency range up to 70 Hz);
- The maintenance of PM is possible in temperature range from absolute zero up to Curie point for piezoelectric materials (i.e. 250 - 300 C0 for the modern materials), and also in vacuum;
- There are no electromagnetic fields and ignitions during the exploitation of one.

According to conducted research PM is complex object for control. Its parameters depend on many factors. Therefore creation of high-precision fast drives based on use of PM is the complicated problem.

To determine the possible directions of drive performances improving based on PM it is important to investigate characteristics of it in details, to analyse different variants of power supply and different control algorithms, to research dynamic characteristics for synthesis of more effective control systems.

3.1. Research of piezomotor as conducting element of drive.

3.1.1. Analysis power supply variants for drive.

PM uses AC voltage. Frequency of it is equal to resonance frequency of PE. Amplitude of power supply is described by the following function: $U \leq U_{max}$ (U_{max} is the maximum voltage of power supply for defined type of PE). The signal can be of any form:

a. Sine wave (sinusoidal);
б. Triangular;
в. Rectangular.

Optimal power supply for PM is the sinusoidal signal. In case of use other signals power of higher harmonics is dispersed on heat of PM.

3.1.2. Research of dynamic properties of drive.

Processes of speed-up and the retardation characterise the dynamics of drive. The experimental research of several types of rotary piezoelectric drives (potency up to 10 W) has been conducted. The rotary speed of PM is measured by tachometer generator during speed-up and retardation.

The experimental research shows that investigated PMs have the similar dynamic characteristics.

The processes of speed-up and retardation of the drive PM-28 are shown at Fig 1.

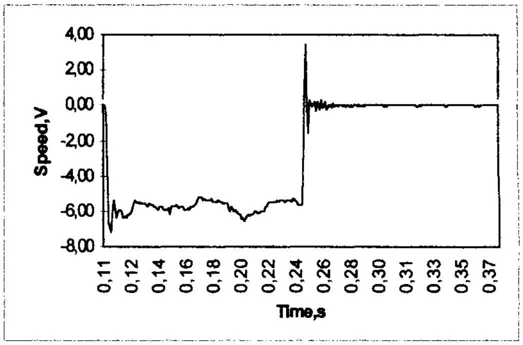


Figure. 1.

The analysis of processes has shown, that:

- The transient process is finished after some milliseconds;
- Processes of speed-up and the retardation have an expressed oscillatory character;
- The process of speed-up differs from the process of retardation.

Review of the literature has shown that there is no mathematical model enough completely describing PM at present time.

Due to obtained experimental results, at preliminary design stages the piezoelectric drive can be described by an oscillatory unit in the first approximation. Input is the amplitude of power supply signal. Output is the rotary velocity of PM.

The parameters of unit is accepted the following values for PM-28 motor:

$K_{PM} = 13.18$ (degree/sec)/V;
$T_{PM} = 6.3*10^{-4}$ sec;
$\xi = 0.158$

Simulation results of the speed-up and retardation processes for the PM-28 motor by means of PDS program developed at BMSTU are shown at Fig. 2.

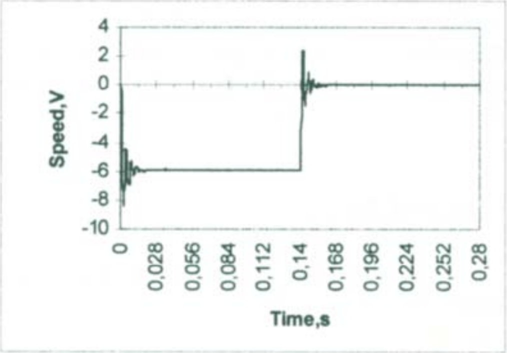


Figure. 2

3.1.3. Analysis of control algorithms for piezomotor.

3.1.3.1. Control by rotary speed

The rotary speed of PM can be controlled by means of change of amplitude, frequency and relative pulse duration [4],[5].[6].

Dependence of rotary velocity from the amplitude of power supply [5] is shown at Fig. 3. Dependence of PM velocity from the power supply amplitude varies under the non-linear law. The velocity of motor depends on load moment at the shaft of motor non-linear.

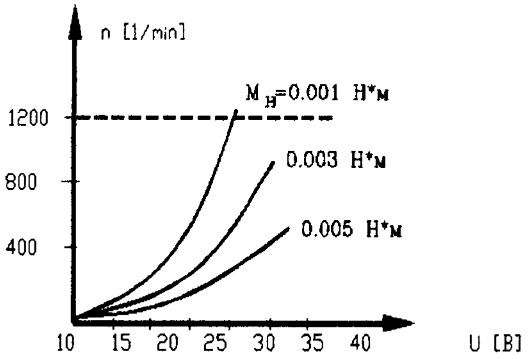


Figure 3.

Dependence of rotary velocity from the frequency of power supply [5] is shown at Fig. 4.

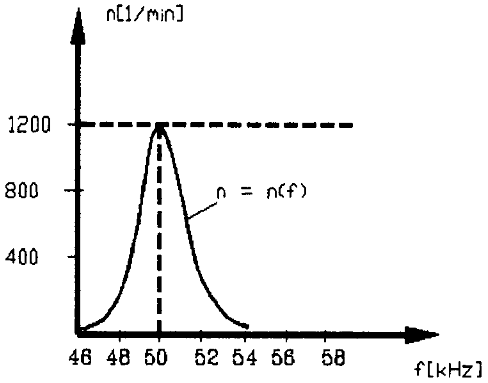


Figure 4.

According to analysis of this Fig. it is clear, that to achieve the best performances it is necessary that the frequency of power supply is equal to the resonance of motor.

Dependence of PM rotary velocity from the relative pulse duration of power supply is drawn at Fig. 5. This dependence is shown for DPER-35K1-1 motor. The dependence type of PM velocity from relative pulse duration is approximately linear.

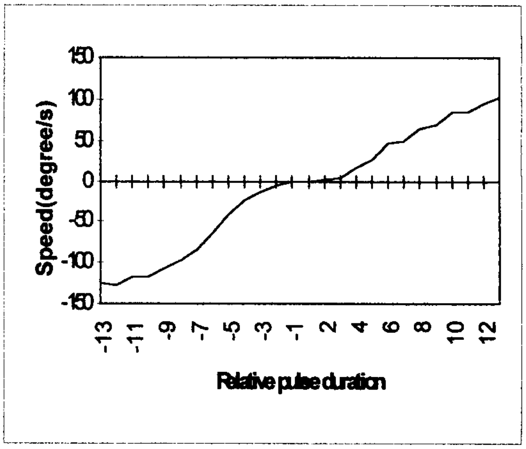


Figure. 5.

Conducted research of stability definition for motor shaft velocity shows the instability of the motor rotary velocity has 5% range.

3.1.3.2. Control by position of motor shaft

The shaft turns after power supply to motor. In case of absence of power supply the shaft is braking. The transient process is finished after some milliseconds, i.e. it follows rather fast. Therefore, it is possible to control with the motor in a start-stop condition effectively.

The resolving ability of shaft positioning is determined by the turn magnitude of the motor rotor during one oscillation of PE. It is called step of a motor. The magnitude of step belongs of several angular seconds range. The corresponding control of the motor allows to positioning the shaft of motor with high accuracy (with accuracy up to step).

However, the driving of PE is transformed to rotation of the rotor via moment of friction. Research of stability determination for PM step magnitude is conducted. Due to experimental research necessity to apply the feedback sensor under position is concluded for fine positioning of the motor shaft.

3.2. Results

Due to conducted research the control algorithms for PM are created, which allow to design precision high dynamic without reducer drives based on piezomotors.

4. CONCLUSION

With the purpose of perfection PA expediently to carry out researches opportunity to use of same build-in sensors for correction of its dynamic properties.

At present time, new more effective control algorithms for PM is developed. It permits to reduce the influence of the exterior factors (for example, temperature) on work of a drive completely.

REFERENCES

1. Issartel, J.-P., Multilayer piezoelectric actuator stack and method for its manufacture, European Patent Application, No. 0427901A1, 22.05.91 Bulletin 91/21, European Patent Office, 1991.

2. Kirimlioglu S.Savas, Anlagan Omer, Kilic S.Engin, Development of a piezoelectric actuator to be used in piston manufacturing, 7th International Machine Design and Production Conference, September 11-13, 1996, METU, Ankara, Turkey.

2. Müller F, Piezoelektrische Stellatriebe für lineare Bewegungen, Feingeratetechnik, Berlin 35 (1986) 11.

4. Vishnevsky V.S., Kartashev I.A., Lavrinenko V.V. Piezoelectric drives. - M.:Energia, 1980.- 112 p.

5. Piezoelectric drive. /V.V. Lavrinenko, V.S. Vishnevsky, O.L.Boychenko etc. - Electricity, 1981, № 6, pp. 68-70.

6. Erofeev A.A. Piezoelectric devices of an automation. - L.: Mashinostroenie, 1982. - 212 p.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

An SMA-Driven Micropump Using Electro-Conjugate Fluid Jet Cooling

Shinichi YOKOTA[a], Kazuhiro YOSHIDA[a], Tomohide KAWAUE[b], Yasufumi OTSUBO[c] and Kazuya EDAMURA[d]

[a] Precision and Intelligence Laboratory, Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, JAPAN

[b] Graduate School, Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, JAPAN

[c] Center of Cooperative Research, Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba 263-0022, JAPAN

[d] New Technology Management Co., Ltd.
2-9-1-306 Higashi-shinkoiwa, Katsushika-ku, Tokyo 124-0023, JAPAN

For high power micromachines using fluid power, micropumps are needed as power sources. A shape memory alloy (SMA) is one of promising actuators for micropumps, however, the response is insufficient on a cooling phase. To overcome the problem, in this paper, an SMA-driven micropump using Electro-Conjugate Fluid (ECF) jet cooling is proposed. ECF is a new kind of functional fluid which produces jet flows between high DC voltage-applied rod type electrodes and can realize simple cooling mechanisms. Then, the basic characteristics of the SMA actuator using ECF cooling is examined experimentally. Finally, a micropump with 15mm in diameter is fabricated and the output characteristics are experimentally investigated. The maximum output flowrate of 13mm^3/s, the maximum pressure of 0.17MPa and the maximum output power of 0.12mW are realized.

1. INTRODUCTION

Research and development of micromachines have been one of active fields of engineering [1, 2]. A maintenance micro system which inspects and repairs the inner wall of small diameter pipes of power plants with about 10mm in diameter is an example of the targets [2]. For such practical micromachines, it is significant to produce enough power to work. Fluid power features high output power density and has advantages for millimeter-size actuators [3]. The authors have been studying on micromachines using fluid power [3, 4]. For fluid micromachines, micropumps are needed as power sources. However, the formerly fabricated micropumps are to transport some liquids or gases and the output power is insufficient for power sources.

A shape memory alloy (abbreviated as "SMA", hereafter) is a promising actuator for a high-power micropump because of the simple structure and the high energy density [3], however, the response is insufficient on a cooling phase. The response determines the pumping frequency and the output flowrate. Two of the authors have realized a bandwidth of 60Hz of SMA actuators for hydraulic proportional control valves by using forced-oil cooling and so on [5], however, the cooling apparatus is difficult to apply for micropumps because of the large size. To overcome the problem, in this study, a new active cooling method using an Electro-Conjugate Fluid (abbreviated as "ECF", hereafter) is proposed. ECF is a new kind of dielectric functional fluid which produces jet flows between high DC voltage-applied electrodes [6]. One of the ECF is dibutyl decanedioate (abbreviated as "DBD", hereafter). By using ECF, a simple liquid cooling mechanism for an SMA actuator is realized. An SMA-driven micropump using the proposed cooling mechanism is expected to produce high output flowrate

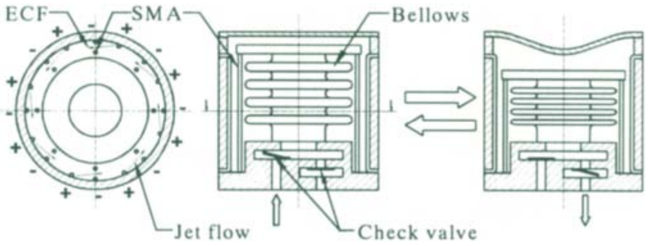


Figure 1. Proposed micropump

irrespective of the viscosity of working fluid.

In this paper, firstly, an SMA-driven micropump using ECF cooling is proposed. Secondly, basic experiments on the SMA actuator using ECF jet cooling is performed. Finally, a micropump with 15mm in diameter is fabricated and the output characteristics are experimentally investigated.

2. PROPOSITION OF SMA-DRIVEN MICROPUMP USING ECF JET COOLING

2.1. Structure of the proposed micropump

The proposed micropump consists of a bellows as a variable chamber, SMA wires, two check valves, and an outside cylinder with rod electrodes for ECF as shown in Fig.1. The SMA wires are connected to the bellows and control the height. Two check valves are located at the bottom of the bellows. One is for the outlet port and the other is for the inlet port. The bellows and the outside cylinder are coaxial and in the space between them, the SMA wires, electrodes for ECF and ECF are placed. The SMA wires are contracted at high temperature with applied current or expanded at low temperature with cooling by ECF jet flows. When the SMA wires are contracted, inside volume of the bellows is decreased and working fluid flows out through the outlet check valve. When the SMA wires are expanded, on the other hand, the inside volume is increased and working fluid flows in through the inlet check valve. By iterating these two movements, working fluid flows out intermittently.

For high output flowrate, high response of the SMA wires is required for high pumping frequency. The contracting response of SMA is able to be improved by increasing the applied current, however, the expanding response is limited by the thermal response which is insufficient in usual cases. In the proposed micropump, on the cooling phase, the heat of SMA wires is transported to the outside cylinder by the ECF jet flows and then radiated through large surface area of the outside cylinder.

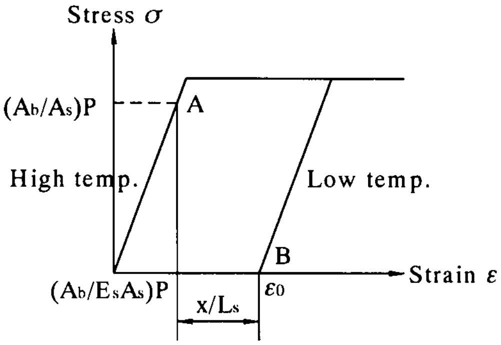


Figure 2. Static characteristics of SMA wire

2.2.Theoretical analysis of the output characteristics

A relation between stress and strain of an SMA wire is illustrated in Fig.2. In the proposed micropump, the point A is corresponding to the contracted state after flowing out with high load force. The point B is corresponding to the expanded state after flowing in with no-load force. Hence the flowrate Q for output pressure P is calculated as follows:

$$Q = A_b x f = Q_{\max}\left\{1-\left(P/P_{\max}\right)\right\}, \qquad (1)$$

where,

$$\begin{cases} Q_{\max} = A_b L_s \varepsilon_0 f \\ P_{\max} = \varepsilon_0 E_s A_s / A_b \end{cases},$$

and x, f, L_s, A_s, E_s are amplitude, driving pulse frequency, length, total sectional area and Young's modulus of the SMA wires, respectively, and A_b is effective sectional area of the bellows. The output characteristics are represented as a line connecting two points (P, Q)=(0, $Q_{\max}$) and ($P_{\max}$, 0). In the proposed micropump, the driving pulse frequency which produces the same temperature change is increased by the ECF jet cooling method and the output flowrate is increased.

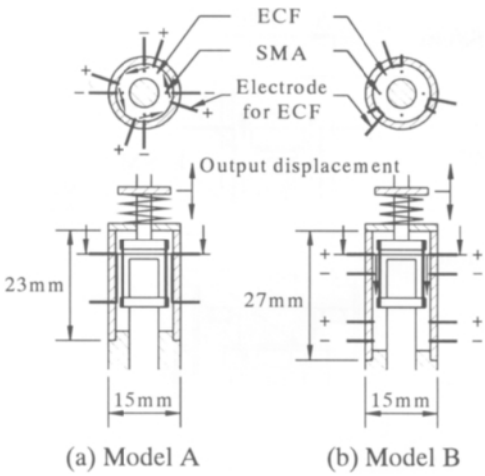


Figure 3. Fabricated SMA actuation units

3. EXPERIMENTS ON SMA ACTUATOR USING ECF JET COOLING

3.1. Experimental apparatus

The basic characteristics of the SMA actuator using ECF jet cooling are experimentally investigated. The fabricated actuation units are shown in Fig.3. Each actuation unit consists of a linearly movable shaft with a small spring, SMA wires, and a cylinder with electrodes for ECF. The shaft is moved by the SMA wires. The SMA wires are heated by Joule's heat and cooled down by the ECF jet flows. To examine suitable arrangement of the electrodes for ECF, two models called "the model A" shown in Fig.3(a) and "the model B" shown in Fig.3(b) are fabricated. In both models, the electrodes are rod type and located on the inside wall of the cylinders. The model A has rod electrodes parallel to the cylinder axis and produces circulation flows along the inside wall of cylinder. The model B has rod electrodes vertical to the cylinder axis and produces circulation flows along the cylinder axis. The SMA wires are Ti-50Ni (at%) wires with 75μm in diameter, 14mm in length and the number is 4. In the experiments, periodic pulse current is applied to the SMA wires and the output displacement of the shaft is measured by a linear potentiometer in the steady state. The pulse width is 20ms and the steady state is judged when the time variation of the amplitude is less than 3%. The ECF is DBD. Model types, applied voltage for ECF, applied power and driving pulse frequency for SMA are varied and the amplitude of SMA wires are evaluated.

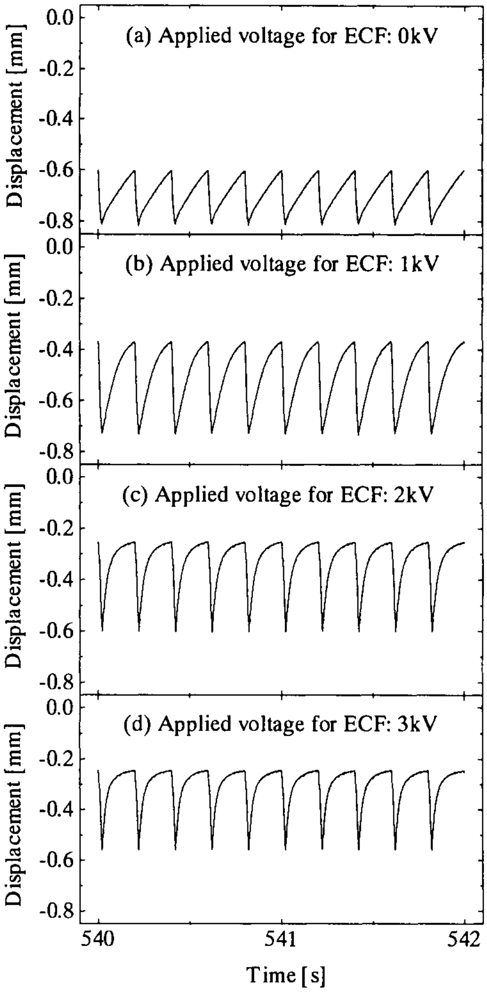
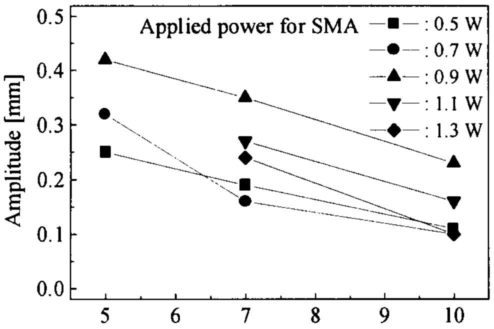


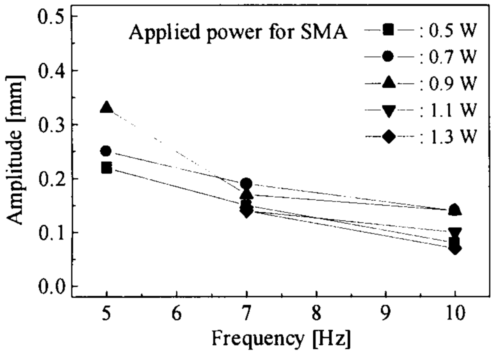
Figure 4. Examples of measured output displacement of the SMA actuation unit (Model A, applied power for SMA: 0.7W, driving pulse frequency: 5Hz)

3.2. Experimental results

Figure 4 shows examples of the measured output displacement of the model A at driving pulse frequency of 5Hz. When current is applied, the output



(a) Output amplitude of the model A



(b) Output amplitude of the model B

Figure 5. Experimental results of the actuation units (Applied voltage for ECF: 3 kV)

displacement is decreased rapidly and when the current is cut off, the output displacement is recovered. Without voltage applied for ECF, as the average temperature of SMA is high, the minimum displacement and the amplitude are small as shown in Fig.4(a). With larger voltage applied for ECF, as the cooling performance is improved, the minimum displacement and the amplitude become large as shown in Figs.4(b) and (c). The effectiveness of the ECF cooling is also indicated in the waveforms of the output displacement. With too high voltage applid for ECF, the minimum displacement becomes large, however, as the cooling performance is too much, the phase transformation of SMA is not occurred enough and the amplitude becomes small as shown in Fig.4(d). For the model B, similar characteristics are obtained.

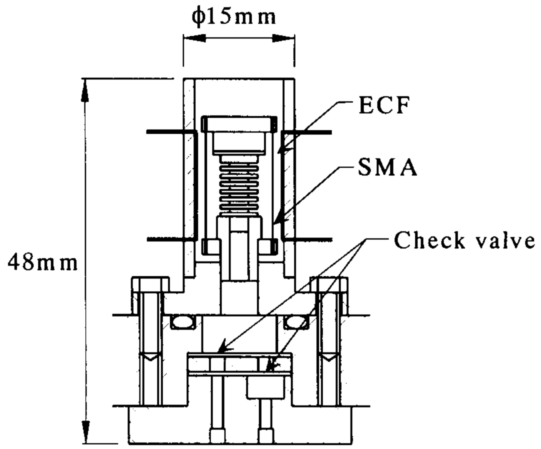


Figure 6. Fabricated micropump

Figure 5 shows the relation between the amplitude and the driving pulse frequency when 3kV is applied for ECF. It is found that the amplitude for the model A is larger than that for the model B and the cooling performance of the model A is higher than the model B. It is also found that the amplitude has the largest value when 0.9W is applied for the SMA.

As a result, it is clarified through these experiments that the best conditions for the SMA actuator are the model A, the applied voltage of 3kV for ECF and the applied power of 0.9W for SMA.

4. FABRICATION OF MICROPUMP AND ITS OUTPUT CHARACTERISTICS

4.1. Fabrication of a micropump

Figure 6 shows the fabricated micropump with 15mm in diameter, 48mm in height based on the results mentioned in the previous section. Length and the number of the SMA wires are 12mm and 4, respectively. The total electric resistance is 13Ω. The check valves are constructed with a rubber sheet with 0.2 mm in thickness. The working fluid is tap water with the viscosity of 1mPas. To investigate the basic characteristics of the proposed pump, the size of the check valves are not so small, however, with eliminating some dead volume, the pump size will be 15mm in diameter and 30mm in height.

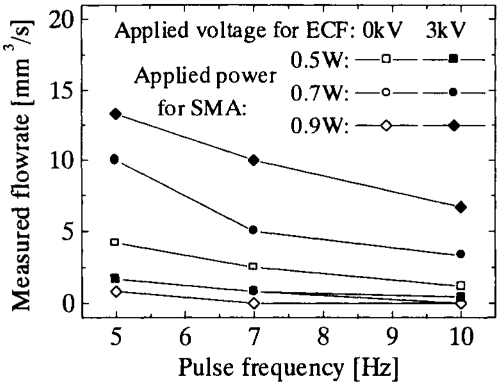


Figure 7. Measured flowrate with no-load

4.2. Output flowrate with no-load

The output flowrate with no-load pressure of the fabricated micropump is experimentally examined. Flowrate is calculated with outflow volume measured by a measuring cylinder in 2min.

The experimental results are shown in Fig.7. To prevent overheat of the SMA, in the experiments, the lowest driving pulse frequency is limited to 5Hz.

As a result, the maximum flowrate of 13mm³/s is realized with the applied voltage of 3kV for ECF, the applied power of 0.9W and the driving pulse frequency of 5Hz for SMA. The value is 57% of the value calculated from the measured SMA amplitude in the previous experiment, the driving pulse frequency and the effective sectional area of the bellows. The loss is thought to be caused by residual bubbles in the bellows, the limitation of bellows displacement and so on.

4.3. Output characteristics

As a pumping performance, the output flowrate with applied load pressure of the fabricated micropump is measured by using the experimental apparatus shown in Fig.8. As the output flowrate is small and it is difficult to fabricate an adequate size orifice, in this study, a bellows is attached to the outlet as a load. The load pressure is increased with time process. The inflow Q is calculated as follows:

$$Q = A_{bL}\frac{dx_L}{dt}, \tag{2}$$

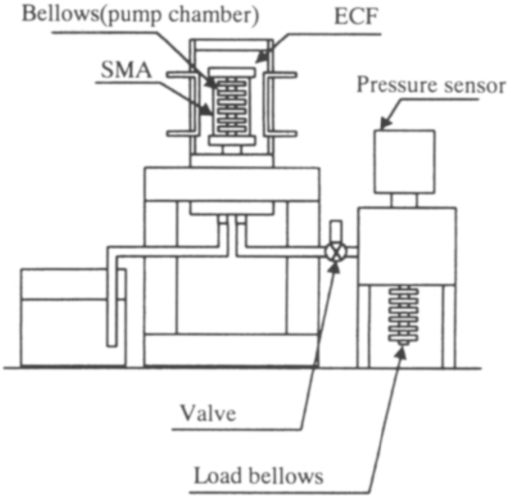


Figure 8. Experimental appratus with load

where, A_{bL} and x_L are the effective sectional area and the displacement of the load bellows. Without external force for the load bellows, the displacement x_L is calculated as follows:

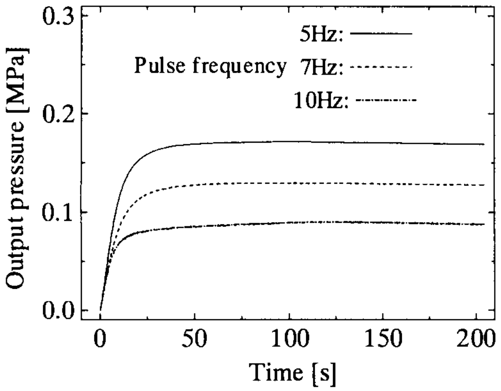
$$x_L = \frac{A_{bL}P}{K_{bL}}, \tag{3}$$

where, K_{bL} is the spring constant of the load bellows. Substituting equation (3) for equation (2), the flowrate Q is calculated as follows:

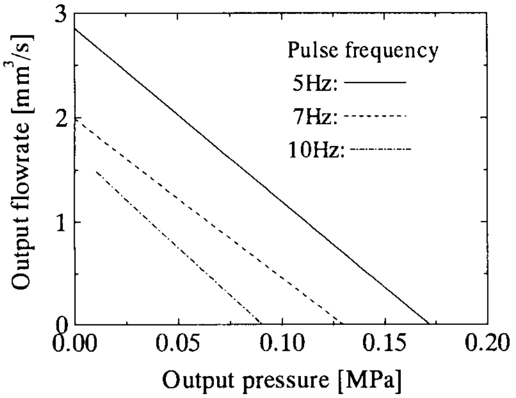
$$Q = \frac{A_{bL}^2}{K_{bL}}\frac{dP}{dt}. \tag{4}$$

In the experiments, the output flowrate Q is calculated from the measured output pressure P based on the equation (4). The effective sectional area, the length and the spring constant of the load bellows are A_{bL}=37mm², L_{bL}=13mm, and K_{bL}=8.8N/mm, respectively. The sampling frequency is 10Hz in the experiments.

Figure 9(a) shows the time variation of the output pressure of the micropump. Based on the Fig. 9(a), it is found that the response of the output pressure is time-lag of first order. This means that the output flowrate is decreased when the output pressure is increased. It is also found that the maximum output pressure is



(a) Pressure variations



(b) Relation between flowrate and pressure

Figure 9. Measured output characteristics with load

decreased with larger driving pulse frequency. This is thought to be caused by the residual bubbles, because when the driving pulse frequency is large, the amplitude of the SMA is decreased as shown in Fig.5 and the affection of the residual bubbles becomes large. It is clarified that the maximum output pressure P_{max}=0.17MPa is realized.

Figure 9(b) shows the relation between the output pressure and the output flowrate calculated from Fig. 9(a). A linear relation which is derived in section 2.2 is verified, although the maximum output flowrate with no-load pressure Q_{max}=2.8mm^3/s is less than the measured value in section 4.2. Based on the results, the maximum output power is 0.12mW.

5. CONCLUSIONS

In this paper, as a power source for practical micromachines, an SMA-driven micropump using ECF jet cooling is proposed and the performance is experimentally examined. As a result, the effectiveness of the ECF jet cooling is verified and the best driving conditions are experimentally derived. Also, a micropump with 15mm in diameter is fabricated and the maximum flowrate of 13mm^3/s, the maximum pressure of 0.17 MPa and the maximum output power of 0.12mW are realized.

REFERENCES

1. Proc. IEEE 10th Annual Int. Workshop on Micro Electro Mechanical Systems (MEMS '97), Nagoya, Japan, (1997).
2. Y.Tatsue, National R & D Program for Micro Machine Technologies, Vol. 12, No. 4, (1994) 508 (in Japanese).
3. K.Yoshida and S.Yokota, Study on High-Power Micro-Actuator Using Fluid Power, Preprints 6th Int. Conf. on Flow Measurement (FLOMEKO '93), Seoul, Korea, (1993) 122.
4. K.Yoshida, H.Mawatari, S.Yokota, An In-Pipe Mobile Micromachine Using Fluid Power Traversable Branched Pipes", Proc. 3rd JHPS Int. Symp. on Fluid Power Yokohama '96, Yokohama, Japan, (1996) 229.
5. S.Yokota, K.Yoshida, K.Bandoh and M.Suhara, Response of Proportional Valve Using Shape-Memory-Alloy Array Actuators, Preprints 13th World Congress of IFAC, San Francisco, USA, Vol.A (1996) 505.
6. S.Yokota, K.Yoshida, Y.Otsubo and K.Edamura, A Micromotor Using a Kind of Dielectric Fluids, Conference Record 1996 IEEE Industry Applications Society 31st IAS Annual Meeting, San Diego, USA, Vol. 3 (1996) 1749.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Information Transmitting Properties of Mechatronic Systems

Kaposvári Z., Rostás J., Szabó T.

Technical University of Budapest, Department of Precision Engineering and Optics
H-1521 Budapest, Egry J. u. 1. HUNGARY

For certain mechatronic systems (e.g.: measuring instruments and some control equipment) the emphasis is often put on accuracy of information acquisition and transmitting while the type and magnitude of the physical quantity carrying the information is of secondary importance. In this paper have been elaborated by the authors a method and a characteristic number for determining of the information transmission efficiency of mechatronical systems.

1. INTRODUCTION

In our time the concepts of the information, as well as the connected processes, are very important and well known. The usefulness and efficiency of the concept applicable theory caused the interests forward him. Nowadays the different special sciences are utilising the results of the information theory. Simultaneously the connected uncertainty theories are developing rapidly.

Recognising the advantages of this concept we are also trying to use out the results of the theory for qualifying the different mechatronical systems.

In this lecture we elaborated a method for the investigation of, so named, transfer function and we demonstrate the application of it, for two different systems.

2. THE ENTROPY OF THE TRANSFER FUNCTION

A given input-output system can be view by the scheme of Fig.1

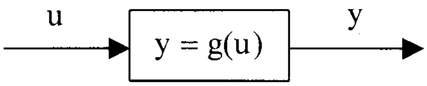


Fig.1

Where $u = (u_1, u_2, \dots, u_n)$ demonstrates the vector of input variables and $y = (y_1, y_2, \dots, y_n)$ the vector of output. It is assumed, that the transfer function of the system is known:

$$y = g(u) \qquad (1)$$

We can view the uncertainty of the y variable through the entropy:

$$H(y) = -\int h(y) \ln(y) dy \qquad (2)$$

Where $h(y)$ is the joint density function of the y variables. Through the u variables can we express in the form:

$$H(y) = H(u) + \int f(u) \ln \frac{\partial g}{\partial u} du \qquad (3)$$

Where

$$J = \left| \frac{\partial g}{\partial u} \right| \qquad (4)$$

is the Jacobean of the function g. Otherwise the entropy of the y:

$$H(y) = H(g(u)) = H(g, u) \qquad (5)$$

While it is known, that:

$$H(g, u) = H(u) + H(g|\ u) \qquad (6)$$

Comparing (3), (5) and (6) we can obtain:

$$H(g|\ u) = \int f(u) \ln \frac{\partial g}{\partial u} du \qquad (7)$$

Where $f(u)$ is the joint density function of the u variables.

The equation (3) has two consequences:

- If the Jacobean determinant of function g equal to the unity, so the second term of the expression (3) vanishes, the entropy of $H(y)$ and $H(u)$ will be equals. I.e. the u variables will be observable through the variables y. The observation will be entropy-invariant.
- If the value of the Jacobean determinant deviates from the unity, so the value of the second term gives information for the quality of the transfer function (for the correctness of the mathematical model).

3. THE QUALIFICATION OF THE TRANSFER FUNCTION

For the mutual transferred information trough the function $g(u)$ introduced the term:

$$I_g(I; u) = H(y) - H(g|\ u) \qquad (8)$$

one can define the efficiency of the function g as follows:

$$\mu_g = \frac{I_g(y; u)}{H(y)} \qquad (9)$$

or rather after eq. (8)

$$\mu_g = 1 - \frac{H(g|\ u)}{H(y)} \qquad (10)$$

On basis of eq. (3) $\quad \mu_{y|u} + \mu_{g|u} = 1 \qquad (11)$

where

$$\mu_{y|u} = \frac{H(u)}{H(y)} \qquad (12)$$

$$\mu_{g|u} = \frac{H(g|\ u)}{H(y)} \qquad (13)$$

Taking these into consideration:

$$\mu_g = 1 - \mu_{g|u} \qquad (14)$$

These expressions show the sharing of the transfer efficiency and the uncertainties.

4. THE TRANSFER FUNCTION OF THE OPTIMETER

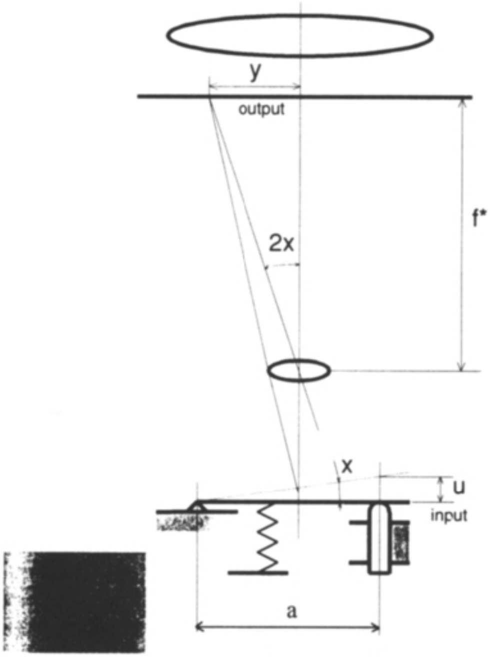


Fig.2

On basis of the Fig.2 one can write:

$$y = \mathrm{tg}\, 2\alpha = f^{*} \frac{2\, \mathrm{tg}\, \alpha}{1 - \mathrm{tg}^{2} \alpha} \tag{15}$$

$$\mathrm{tg}\, \alpha = \frac{u}{a} \tag{16}$$

The transfer function will be:

$$g(u) = 2f^{*}a \frac{u}{a^{2} - u^{2}} \tag{17}$$

Then the uncertainty of the transfer function (Appendix 1):

$$H(g|\ u) = \ln\left(2 \frac{f^{*}}{a}\right) + 3 \frac{\alpha_2}{a^2} + \frac{\alpha_4}{2a^4} \tag{18}$$

or rather in terms of expected value (μ_u) and of the standard deviation will be as follows (19)

$$H(g|\ u) = \ln\left(2 \frac{f^{*}}{a}\right) + \left\{3 \frac{\mu_u}{a^2} + 3 \frac{\sigma_u}{a^2} + \frac{\alpha_4}{2a^4}\right\}$$

Because $u = 200\ \mu m$
$a = 5\ mm$
$f^{*} = 200\ mm$

have values mentioned above, the quantity of the second term in eq.19 is negligible respect of the first one, except in the case if:

$$\frac{f^{*}}{a} = \frac{1}{2} \tag{20}$$

then

$$\ln\left(2 \frac{f^{*}}{a}\right) = 0 \tag{21}$$

so the uncertainty should be depended only on the second term. In the example of the mentioned optimeter this case is not realisable.

5. THE TRANSFER FUNCTION OF PIEZOELECTRIC ACTUATOR

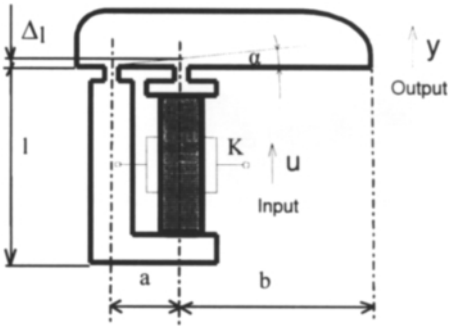


Fig.3

On basis of the Fig.3 we can write:

$$\operatorname{tg}\alpha = \frac{\Delta l}{a} \text{ and } \operatorname{tg}\alpha = \frac{y}{a+b} \qquad (22)$$

$$g(u) = y = (a+b)\operatorname{tg}\alpha = (a+b)\frac{\Delta l}{a} \qquad (23)$$

Assuring a linear relation between the voltage connected to the actuator (u) and the translation (Δl) causing through the same the voltage as follows:

$$\Delta l = Ku \qquad (24)$$

Where K the characteristic constant of the actuator.

The transfer function of the actuator (Appendix 2)

$$g(u) = \left(1 + \frac{b}{a}\right)Ku \qquad (25)$$

Then the uncertainty of the transfer function:

$$H(g|\ u) = \ln\left\{\left(1 + \frac{b}{a}\right)K\right\} \qquad (26)$$

This uncertainty is the smallest, if

$$\left(1 + \frac{b}{a}\right)K = 1 \qquad (27)$$

that is to given K you have to choice such a $\frac{b}{a}$ ratio, where

$$\frac{b}{a} = \frac{1}{K} - 1 \qquad (28)$$

6. DISCUSSION OF RESULTS

ad1. The calculation of the information transmission for the optimeter.
Assuming that the output normally distributed, so according to [2] we get:

$$H(y) = \frac{1}{2}\ln\left(2\pi e\sigma_y\right) + \ln\Delta y \qquad (29)$$

where the standard deviation of the output:
$\sigma_y = 1000\ \mu m$
the division of scale: $\Delta y = 1000\ \mu m$

according to eq. (28) $H(y) = 11.78$
according to eq. (19) $H(g|u) = 4.87$
according to eq. (10) $\mu_g = 0.63$

On the basis of calculation we can determine that the information transmission efficiency of the optimeter is: 63%

ad2. The determination of the information transmission efficiency of the piezoelectric actuator:
Assuming that the output is normally distributed, as well as the standard deviation of the output signal is: 100 nm
the magnification is: $\frac{b}{a} = 2$
the resolution: $\Delta y = 60$ nm
the sensitivity: $K = 5$ nm/V

according to eq. (28) : $H(y)=7.81$
according to eq. (25) : $H(g|u) = 2.78$
according to eq. (10) : $\mu_g = 0.65$

In case of the given parameters the information transmission efficiency of the transducer is: 65%

The presented examples are demonstrating, that, the information transmission of the investigated mechatronic systems can be influenced first by the varying of geometric (design) parameters, secondly by the making better of the mechatronic model i.e. the transfer function.

APPENDIX 1

Substituting to the (7) form of the entropy:

$$H(g|\ u) = \int f(u)\ln\frac{\partial g}{\partial u}du \tag{1.1}$$

$$\frac{\partial g}{\partial u} = 2f^{*}a\frac{a^{2}-u^{2}+2uu}{\left(a^{2}-u^{2}\right)^{2}} \tag{1.2}$$

$$\frac{\partial g}{\partial u} = 2f^{*}a\frac{a^{2}+u^{2}}{\left(a^{2}-u^{2}\right)^{2}} \tag{1.3}$$

$$\ln\frac{\partial g}{\partial u} = \ln 2f^{*}a\frac{a^{2}+u^{2}}{\left(a^{2}-u^{2}\right)^{2}} = \tag{1.4}$$

$$= \ln(2f^{*}a) + \ln(a^{2}+u^{2}) - 2\ln(a^{2}-u^{2})$$

Substituting to the (1.1) form:

$$\begin{aligned} H(g|\ u) &= \int\ln(2af^{*})f(u)du + \\ &+ \int\ln(a^{2}+u^{2})f(u)du - \\ &- 2\int\ln(a^{2}-u^{2})f(u)du \end{aligned} \tag{1.5}$$

Taking into account the feature of the density function i.e.:

$$\int_{+\infty}^{-\infty} f(u)du = 1 \tag{1.6}$$

Computing the integrals separately:

$$\int\ln(2af^{*})f(u)du = \ln(2af^{*}) \tag{1.7}$$

$$\begin{aligned} &\int\ln(a^{2}+u^{2})f(u)du = \\ &= 2\ln a + \int\ln\left(1+\frac{u^{2}}{a^{2}}\right)f(u)du = \\ &= 2\ln a + \int\left(\frac{u^{2}}{a^{2}} - \frac{u^{4}}{2a^{4}}\right)f(u)du = \\ &= 2\ln a + \frac{\alpha_{2}}{a^{2}} - \frac{\alpha_{4}}{2a^{4}} \end{aligned} \tag{1.8}$$

Where

$$\alpha_{i} = \int u^{i}f(u)du \tag{1.9}$$

the i^th^ momentum.

$$\begin{aligned} &\int\ln(a^{2}-u^{2})f(u)du = \\ &= 2\ln a + \int\left(1-\frac{u^{2}}{a^{2}}\right)f(u)du = \\ &= 2\ln a - \frac{\alpha_{2}}{a^{2}} - \frac{\alpha_{4}}{2a^{2}} \end{aligned} \tag{1.10}$$

Substitute the calculated values (1.7), (1.8) and (1.10) into (1.5):

$$\begin{aligned} &H(g|\ u) = \ln(2af^{*}) + \\ &+ 2\ln a + \frac{\alpha_{2}}{a^{2}} - \frac{\alpha_{4}}{2a^{4}} - \\ &- 2\left\{2\ln a - \frac{\alpha_{2}}{a^{2}} - \frac{\alpha_{4}}{2a^{4}}\right\} \end{aligned} \tag{1.11}$$

since

$$\ln(2af^{*}) = \ln(2f^{*}) + \ln a \tag{1.12}$$

after reducing we can get:

$$H(g|\ u) = \ln\left(2\frac{f^*}{a}\right) + 3\frac{\alpha_2}{a^2} + \frac{\alpha_4}{2a^4} \qquad (1.13)$$

APPENDIX 2

The transfer function of the actuator:

$$g(u) = (a + b)\frac{Ku}{a} = \left(1 + \frac{b}{a}\right)Ku \qquad (2.1)$$

Substituting to the (7) form of the entropy:

$$H(g|\ u) = \int f(u)\ln\frac{\partial g}{\partial u}du \qquad (2.2)$$

$$\frac{\partial g}{\partial u} = \left(1 + \frac{b}{a}\right)K \qquad (2.3)$$

Then the uncertainty of the transfer function:

$$H(g|\ u) = \ln\left\{\left(1 + \frac{b}{a}\right)K\right\} \qquad (2.4)$$

PREFERENCES

1. Prof. Tim King Ph.D.,MDesRCA, B.Sc.(Eng), Wei Xu B Eng and Dr John Thornley Ph.D.,B.Sc.: A Piezoelectric Harmonic Motor, Proceedings of the Joint Hungarian - British International Mechatronics Conference (September 21-23.,1994. Budapest), Computational Mechanics Publications, pp.: 527-531.
2. Dr János Rostás, Dr Zoltán Kaposvári, Mihály Szalai: Qualification of Measuring Instruments on the Basis of Information Theory, Periodica Polytechnica, Budapest, Vol.40. No. 2. 1996. pp. 121-142.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

New polarized electromagnetic actuators as integrated mechatronic components - design and application

E. Kallenbach[a], H. Kube[a], K. Feindt[a], V. Zöppig[a], R. Hermann[b], and F. Beyer[b]

[a]Technical University of Ilmenau, Faculty of Mechanical Engineering, Institute for Micro System Technology, P.O. Box 100565, 98684 Ilmenau, Germany

[b]Steinbeis foundation for economic promotion, transfer center mechatronics Ilmenau, Ehrenbergstr. 11, 98693 Ilmenau, Germany

This paper presents an overview and an example of the design process of mechatronic devices using polarized electromagnetic actuators.

1 GENERAL DESIGN PROCESS

Many modern applications in production, automation, and car base on direct drives. They require small powerful drives. Drives using polarized electromagnetic actuators can be used for this purposes.

The magnetic circuit of polarized electromagnetic drives are characterized by a magnetic circuit which consists partially of permanent magnetic materials. Examples for this class of actuators are polarized solenoids, stepper motors, linear drives, and two coordinate drives.

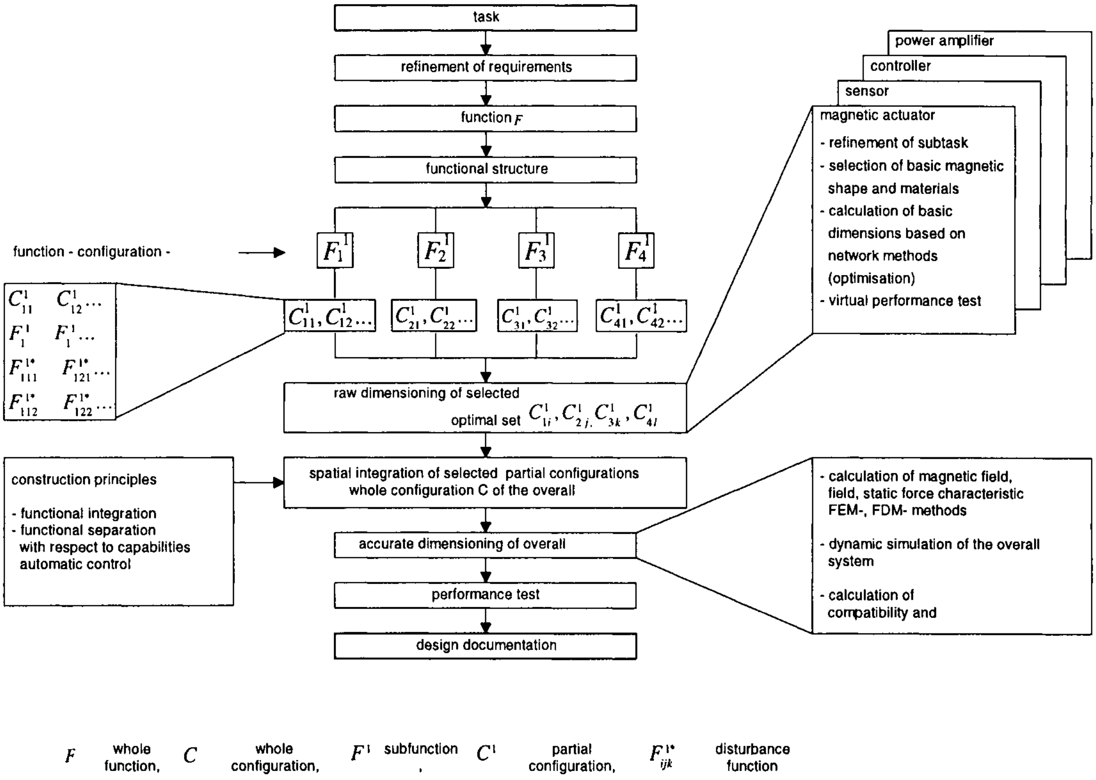

task
refinement of requirements
function F
functional structure
function - configuration -
F_1^1
F_2^1
F_3^1
F_4^1
$C_{11}^1, C_{12}^1 \dots$
$C_{21}^1, C_{22}^1 \dots$
$C_{31}^1, C_{32}^1 \dots$
$C_{41}^1, C_{42}^1 \dots$
$C_{11}^1 \quad C_{12}^1 \dots$
$F_1^1 \quad F_1^1 \dots$
$F_{111}^{1*} \quad F_{121}^{1*} \dots$
$F_{112}^{1*} \quad F_{122}^{1*} \dots$
raw dimensioning of selected optimal set $C_{1i}^1, C_{2j}^1, C_{3k}^1, C_{4l}^1$
power amplifier
controller
sensor
magnetic actuator
- refinement of subtask
- selection of basic magnetic shape and materials
- calculation of basic dimensions based on network methods (optimisation)
- virtual performance test
construction principles
- functional integration
- functional separation with respect to capabilities automatic control
spatial integration of selected partial configurations whole configuration C of the overall
accurate dimensioning of overall
performance test
design documentation
- calculation of magnetic field, field, static force characteristic FEM-, FDM- methods
- dynamic simulation of the overall system
- calculation of compatibility and
F whole function,
C whole configuration,
F^1 subfunction ,
C^1 partial configuration,
F_{ijk}^{1*} disturbance function

Figure 1: design algorithm

They have some advantages in comparison to conventional electromagnetic actuators with neutral magnetic circuit. However, they can generate larger forces, have less power losses, and act faster. Such actuators also allow new drive designs like e.g. linear drives and integrated two co-ordinate drives [1].

Modern high energy permanent magnets made of SmCo and NdFeB have a high coercivity and a high rententivity. Therefore, actuators based on these materials require other design rules than actuators based on classical permanent magnet materials like e.g. barium ferrite and AlNiCo.

Because polarized electromagnetic actuators can also have other behaviors, the demands on the addressing increase, too. For example, bistable polarized actuators need short but powerful current pulses to switch between both end positions.

Therefore, polarized electromagnetic drives are typical mechatronic components. The innovation of these components is that mechanics, sensors, power electronic and information processing interaction.Because the number of design possibilities is large, polarized electromagnetic drives cannot be designed by trial and error. They must be designed in a continuous mechatronic design process. Powerful tools for every single design step are necessary.

Figure 1 shows a possible design algorithm which describes the design principle and the main design steps of polarized electromagnetic drives.

The quality of the design process depends mainly on the structuring of the knowledge about the models and the constructive basic forms which facilitate the step from the function to the corresponding structure [2]. Several models for polarized magnetic circuits are represented and ways for optimization of the actuators are discussed.

In order to find the optimum of the structure, the mechatronic design requires a top-down-strategy. Therefore it is necessary to subdivide the complete function of the drive into basic sub-functions. In order to get a finish configuration, sub-structures are allocated to theses sub-functions.

During the selection of the sub-structures an effective procedure already requires a compatibility check of the functions of the sub-structures. The influence of disturbance the must be also checked and ways to suppress disturbing effects by controlling must be founded.

The design of the whole configuration is the most difficult step. This step is not analyzed enough until today. But, some general rules can be given which are a result of the manufacturing technologies and the related expenses.

For the design of systems manufactured with serial production methods, the principle of the function integration can be recommended for the mechanical part and the principle of function differentiation for the electronic part. Examples of integrated mechatronic drives with polarized electromagnetic actuator are multiple co-ordinate drives [3].

In systems in which mechanic components are manufactured by micro system technologies, the mechanic components may have complex shape, too. The design algorithm (Figure 1) requires multiple models with different abstraction levels for each element and for the whole system.

However, there are no software modules available until today which make a complete CAD of polarized electromagnetic actuators possible and which are be able to communicate. They are only available for neutral electromagnets [4].While powerful development tools are available for the design of control software, meta-models for the design of polarized electromagnetic drive are missed. Selected examples are presented.

Open development systems are required. In order to adapt the development system to an according class of design tasks, the architecture of such a system have to be modular.

To give an example, the design process of a new gripping device is presented. It contains all the different sub-systems of a mechatronic device like actuator, sensor, controller and power amplifier.

2 MINIATURIZED PNEUMATIC VALVE WITH POLARIZED SOLENOID ACTUATOR

There are various directions of the development of miniaturised pneumatic systems. On the one hand, micropneumatics developed for an independent discipline of microsystem engineering. On the other hand, advantages special in automatic control can be achieved with the aid of valves miniaturised strongly.

Pneumatic microsystems consist of integrated pumps, valves, mixers, sensors and so forth. These systems are among other things used in the environmental and in the analytical industry. Both passive and active valves are used. Different driving principles for active valves are known, e.g., electrostatic, piezoelectric or electromagnetic drives. The solenoid principle is characterised by a high energy density and high actuating forces. Therefore, it is the most used principle in conventional pneumatics. However, the three-dimensional character of the magnetic field and the heating due to the Ohm's losses in the coils is problematic for miniaturisation. The energy density can be increased essentially using polarised systems. By the hold in the stable positions without power, the thermal problems decrease.

2.1 Pneumatic valve manufactured by classical technologies

For application in the automation industry, especially the suction gripping devices in automatic assembly stations for the assembly of mini- and microparts a small, powerful, pneumatic valve with low power consumption was needed. Because, such a valve was not commercially available and the available valves does not fulfill the requirements, a new valve has to be developed.

The demand of low power consumption can be fulfilled by using a bi-stable polarized solenoid actuator witch is switched by short current pulses.

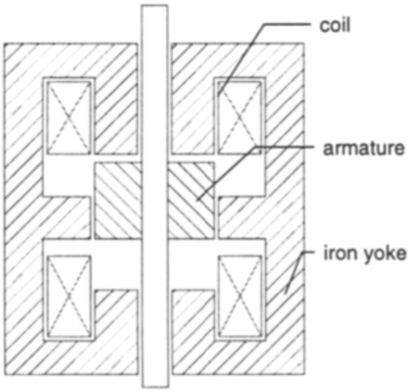


Figure 2 Actuator principle

Figure 2 shows the actuator principle. The armature of the actuators carries a permanent magnetic ring, with generates the forces to hold the armature in the two end positions.

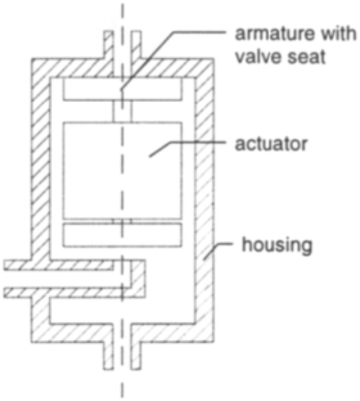


Figure 3 Valve principle

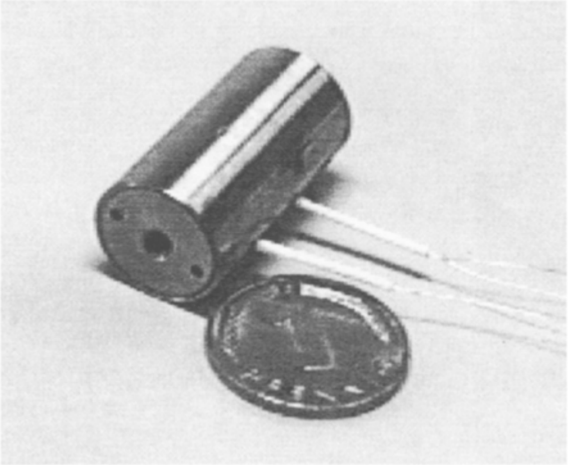


Figure 4 Prototype of the valve

By the use of high-energy compound NdFeB and by the optimization of the magnetic circuit, the size of the actuator could be decreased to a diameter of 10 mm and a length of 10 mm, so that the actuator could be integrated into the valve body (Figure 2).

The valve has a cylindrical shape and a size of 12 mm diameter and 24.5 mm length. Figure 4 shows a prototype of the valve. Inlet and outlet are arranged at opposite sides of the valve with threads to screw in manufacturer specific ports.

The parameters are:

- Bi-stable 3/2-way valve, or 2/2-way valve with integrated drive
- Pressure range: -1 to 8 bar
- Nominal bore: up to 2 mm
- Source voltage: 6 ... 48 Volt (depending on the coil dimensioning)
- Size: Diameter 12 mm, Length 24.5 mm (without ports)

- Weight: app. 14 g (without ports)
- Switching time 1 to 5 ms (depending on the dimensioning, source voltage and working pressure)
- Less power losses (app. 150 mW at 1 Hz working frequency)

2.2 Polarized electromagnetic valve manufactured by modern batch technologies

A second application is a polarized electromagnetic valve using the microstructuring of glass materials manufactured by modern batch technologies. Special material properties like transparency, hardness and chemical resistance allow new applications in medical and analytical instrumentation. Due to the wafer form of the photo-structurable glass material, a layer design of the valves had to be realized. A special high aspect ratio electroplating process is used to form the coils as glass/metal compound structure. These coils support higher current densities than conventional coils.
To implement three dimensional structures without undercut a modified grey scale mask technology is used. The depth of the structures in the glass material is determined by the ratio between width and distance of lines of the mask. The structuring process is shown in Figure 5.

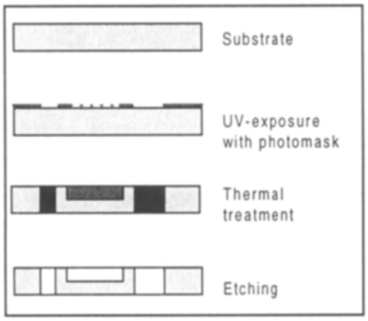


Figure 5 Greyscale mask technology

To implement electromagnetic drives it is necessary to combine the microstructuring technology of glass materials with a high aspect ratio electroplating of copper.
Figure 2 shows the production process of microelectroforming in structured glass.

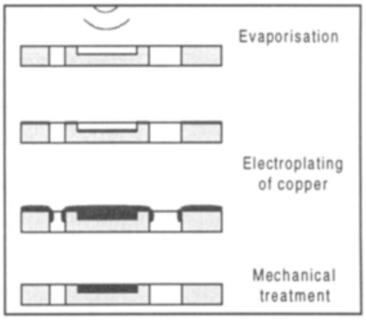


Figure 6 Microelectroforming in structured glass

Figure 7 shows an explosive drawing of the realized microvalve prototype. An SmCo magnet was placed between two iron disks. The bi-stable state can be changed by the switching of short current pulses.
The expected parameters are:

- Valve outlet 0.5 mm
- Force 0.1 N
- Displacement 0.2 mm
- Ampere-turn 30 A

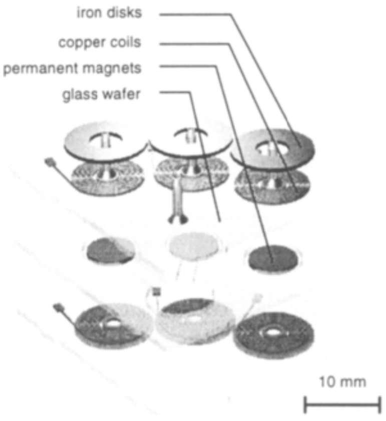


Figure 7 Explosive drawing of a microvalve system in glass

3 DESIGN OF A GRIPPING DEVICE WITH COMPLIANT MECHANISM AND PLANAR MOVING COIL ACTUATOR [10]

A systematic design of grippers requires the analysis of the gripping task which consists of the extraction of features of the parts to be gripped, of the gripping motion, the gripping force dynamics and the gripping periphery [5].
The gripping task is derived from the assembly task. An assembly task often consists of the steps:

- Gripping of the object

- Lifting out of the magazine
- Transport to the place of assembly
- Positioning
- Joining process

For that the grippers have to fulfil the following requirements:

- Secure gripping and releasing performance (problems in particular with releasing due to surface forces and small dimensions),
- guarantee of a small component load by open / closed loop control of the gripping forces and distribution of the gripping forces to as many as possible handle points,
- fastening of the relative position of the gripping object also under action of forces during lifting from the magazine (jamming forces, adhesion forces), transport to the place of assembly (force of inertia), joining process (pressing into fit or into adhesive, forces through shrinkage of adhesives).
- securing of the positioning accuracy through active and passive measures (gripper integrated position sensor, centering structures) and
- system compatibility (clean-room compatibility, application in gripper changing systems, small mass and size).

An essential step during the gripper design is the selection of the gripping principle based on the analysis of the gripping task. Adhesion, pliers and clamping grippers turned out to be especially suitable for the microassembly.

Components produced by microstructuring techniques can effectively fulfill the requirements on grippers for microassembly. Table 1 shows opportunities for the application of microstructured component in the individual subsystems [6].

Table 1 Application of microstructured components

subsystem	applicable microcomponents
actuation system	suction structures, form nests, centering aids, application of structures moulded after the gripping object (positive and negative part)
kinematic system	compliant mechanisms
information processing system	integrated sensors, signal processing and gripper control
drive system	solid body actuators, microactuators

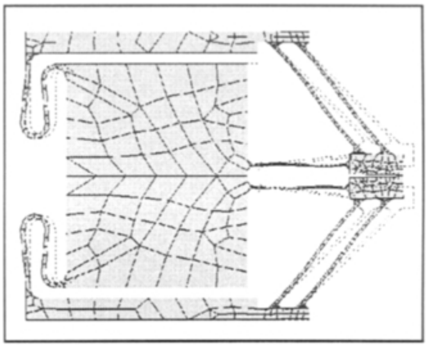


Figure 8 Deformation of the mechanism

Miniature grippers with matter coherent structure and electrodynamic drive fulfill the requirements for the manipulation of microparts [8].

Actuation system: Opposite to speaker systems with cylindrical coils a flat coil was integrated into the gripper as a drive element [7]. Despite the low energy density (0.2 Nmm/cm^3) and low forces the electrodynamic drive has advantages, because of its simple controllability due to its linear force-current relationship (200 mN/A), its great drive displacements in the millimeter range and a very good integration of the flat coil in the compliant mechanism.

The flat coil consists of a glass metal composite (photostructured glass filled with copper by electroplating). It support a higher current density than conventional coils.

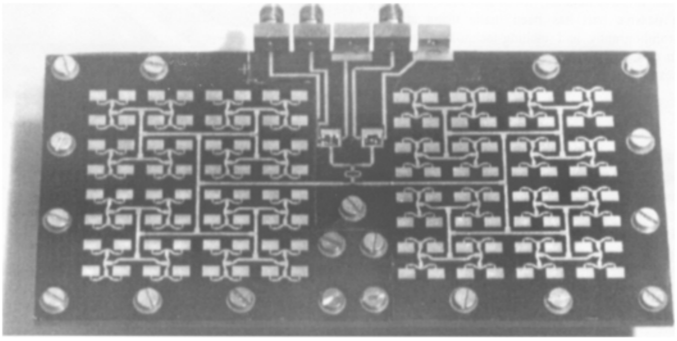


Figure 9 Mechanism with coil

Kinematic system: Because of the great drive displacements a mechanism with low transmission ratio (0.9) could be chosen [8]. Two parallel cranks effect an approximate circular push movement of the gripping jaws. The mechanism was designed with compliant links as deformable parts. With the compliant links an even stress distribution can be achieved and they are suitable for the transmission of great displacements and small forces (electrodynamic principle). Figure 9 shows the mechanism with integrated coil.

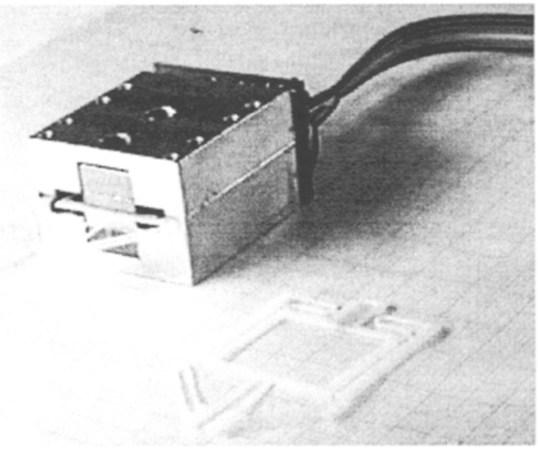


Figure 10 Prototypes of a gripper and a mechanism

The design and calculation of the drive system was done by raw dimensioning using the network method and fine dimensioning using the finite-element-method (ANSYS). Figure 8 shows the deformation of the mechanism due to the Lorentz force.
The mechanism is manufactured by vacuum casting (rapid-prototyping technology) with a special casting resin.
Figure 10 shows prototypes of a gripper and a mechanism.

Information processing system: The linear behavior of both actuation and kinematic system allows to set up a new gripping force measuring principle. In idle mode, i.e. without a gripped object between the gripping elements, there is a defined relationship between coil current I and drive displacement s_A determined by the stiffness of the compliant mechanism (characteristic „idle mode“ in Figure 11). If there is an object between the gripping elements the coil is restricted from further movement as the gripping elements touch the gripped object (point „touching“ in Figure 11). From the coil current I, the difference between drive displacement s_A and drive displacement in idle mode the gripper displacement s_G, gripping force F_G, and the presence of the gripped object can be calculated under consideration of the known transmission properties of the mechanism (stiffness, transmission ration). Figure 11 shows the relationship between drive displacement and coil current or Lorentz force.

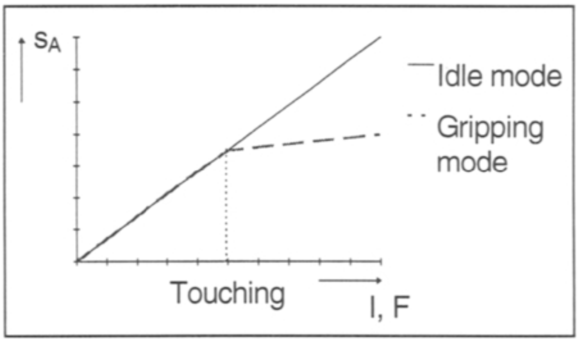


Figure 11 Idle and gripping mode

The measurement of the drive displacement is done with a reflecting opto-coupler realized as integrated SMT-PCB.

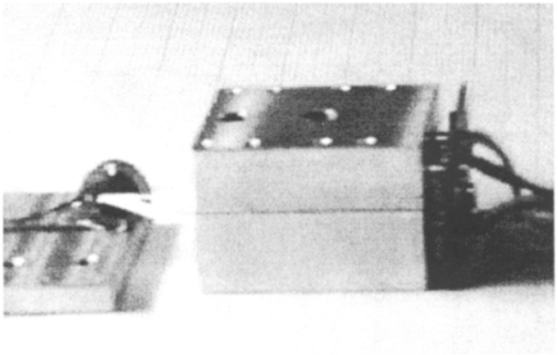


Figure 12 Experimental set-up

In order to calibrate the system an miniaturized force sensor is used as a force sensitive gripping object. Figure 12 shows the experimental set-up with the gripped force sensor.

Figure 13 shows the signals of both gripped force sensor and gripper integrated measurement while excitation with an triangular current waveform (maximum current 1 A, maximum force 70 mN).

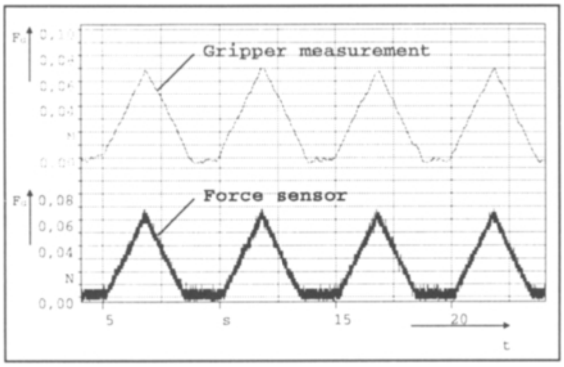


Figure 13 Force signals

The measurement system can be used for a closed loop control (PI-Type) of the gripping force using a PC-I/O-system. Figure 14 shows command variable (rectangular waveform, maximum force 50 mN) and force signal.

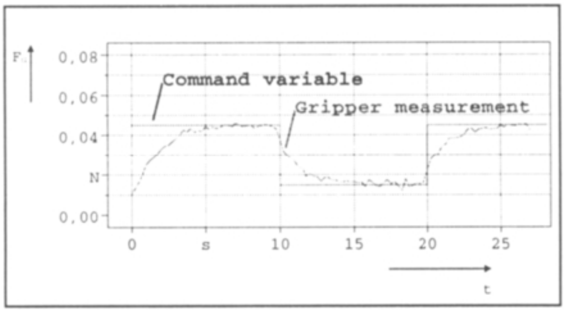


Figure 14 Closed loop control of the gripping force

The closed loop control of the gripping force can be integrated in the SMT-PCB in the next step in order to free the control unit of the assembly station from the control task.

4 REFERENCES

[1] Schäffel, C., Saffert, E., Kallenbach, E.: Integrated Electrodynamic Multicoordinate Drive - Modern Components for Intelligent Motion. ASPE Annual Meeting 1996, Montery, Proceedings p. 456-461

[2] Kallenbach, E.: Configuration of Mechatronic Products with Respect to Control Aspects. VI-CAM Guimaraes Portugal. FPIV Proceedings p. 75-84

[3] Kallenbach, E., Birli, O.: STURGEON - an Existing Software System for the Completely CAD of Electromagnets. ICED 97. Tampere 19-21. August 97, Proceedings vol. 1, p. 147-152

[4] Kallenbach, E., Saffert, E., Birli, O., Räumschüssel, E.: Function Oriented Configuration of Integrated Mechatronic Systems. EFAM Euroconference 21-26. September 1997, Ilmenau, Proceedings p. 107-120

[5] Christen, G.; Pfefferkorn, H.; Gramsch, T.; Zöppig, V.: Flexible Greiftechnik für die Mikromontage. Microengineering 95, Stuttgart 1995

[6] Zöppig, V.; Gramsch, T.; Pfefferkorn, H.; Christen, G.: Miniaturgreifer mit mikrostrukturierten Elementen. Öffentliches Statusseminar im Rahmen der "Mikrotechnik Thüringen '96" des Verbundprojektes "Entwicklung von fertigungsgerechten Montage- und Fügeverfahren". Tagungsband. Jena 1996

[7] Zöppig, V.; Ulrich, J.; Harnisch, A.: Mechanischer Greifer, insbesondere zum Greifen kleiner Objekte. DE 197 15 083 A1

[8] Christen, G.; Pfefferkorn, H.: Greifer mit nachgiebiger Struktur für die Mikromontage. Tagungsband zum 41. Internationalen Wissenschaftlichen Kolloquium der TU Ilmenau, Band 2, 1996

[9] Christen, G.; Pfefferkorn, H.: Linearbewegung mit nachgiebigen Mechanismen. Festschrift zum 65. Geburtstag von Prof. Dr.-Ing. habil. G. Bögelsack. TU Ilmenau, 1997

[10]Christen, G.; Feindt, K.; Gramsch, T.; Harnisch, A.; Pfefferkorn, H.; Zöppig, V.: Microgripping devices with functional elements of microstructured glass materials. Symposium on Handling and Assembly of Microparts. Vienna, November 28-29, 1997

[11]Feindt, K.: Harnisch, A.: Zöppig, V.: Hülsenberg, D.: Kallenbach, E.: 3D-Structring of photosensitive glasses. MEMS 98, Heidelberg, January 25-29, 1998

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

The Recent Advantages of the Axial Flux DC Micromotors

Dr. Attila Halmai

Associate professor, Department Precision Engineering and Optics
Technical University of Budapest, H-1111 Budapest, Egri J. u. 1.

In the Technical University of Budapest, at the Department Precision Engineering and Optics has developed a family of the new axial flux DC micromotors, as a new electromagnetic type actuators for different applications on the field of mechatronics. This DC motors has some advantages with respect the traditional cylindrical, bell-form motors.

1. INTRODUCTION AND THE MAIN PROPERTIES OF THE NEW MOTOR

There are different types of air gap winding (ironless) DC motors. One of those is the well known type, having a bell-formed rotor coil. The other, relatively new designed type of the air-gap winding motors has flat coils. This type of DC motors is called disc motors. They have some advantages with respect to the cylindrical motors, having a bell-formed rotor winding.

First of all: The utilization of the hard-magnetic material is better, than in the case of the cylindrical design. It is an important problem, because the price of the modern rare-earth permanent magnet material is relatively high. If the cost of hard magnetic material was lower, the cost of the complete motor would be also cheaper.

Secondly: The disc-form of the motor gives an excellent possibility to combine this disc motor with the also disc-formed planet-gears, or harmonic drives. This combination of motor and gear-box offers a new direction in the design of robot moving mechanisms.

Thirdly, perhaps the more important feature is as follows: The disc type motors have an extra possibility, which have not the cylindrical ones: to the flat coils can applied an integrated heat-sink, rotating together with the rotor. By this design may achieve better thermal characteristics, namely the thermal constant of the rotor will be more times higher, as one of the traditional motors. The Fig. 1. shows the cross section of a disc motor, having to the coil integrated heat-sink.

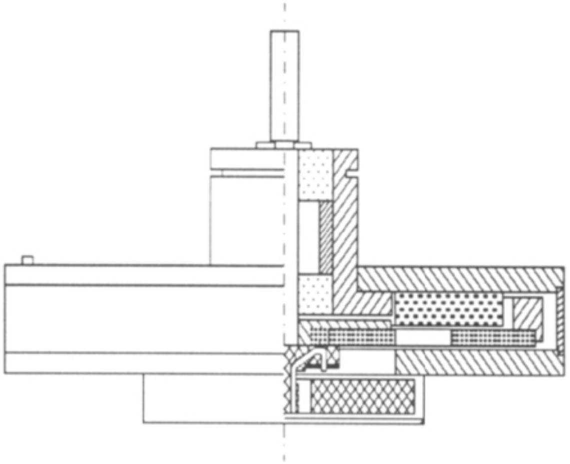


Fig. 1. The cross section of the new axial-flux motor

On the figure are well shown the flat rotor coils, the metallic ring formed heat-sink, the hard and soft magnetic parts, and the commutator. It is advantageous to produce this heat-sink from the material having a good thermal conductivity. It is true, this metallic ring also increases the mechanical inertia of the rotor, but the increase of the thermal constant of the rotor will be approximately ten times greater, than the increase of the mechanical inertia.

At such applications, where the extreme dynamic operation is not needed, the new design gives new perspectives. Namely, the operation area of this motors is more extensive, than the motors without heat-sink. The situation is the same, when a transistor works with or without a heat-sink. The motors applying heat-sink have higher resistance against the short time overloads.

The heat-ring has a further advantage: it gives also a more mechanical stability. If the coils are excited with current, their temperature of it increases, and the mechanical stiffness of the winding decreases. The integrated metallic ring holds on and fixes on the place the coils against the centrifugal forces.

2. THE ELECTROMECHANICAL AND THERMAL PROPERTIES OF THE NEW MOTOR

In the next table is shown the main electromechanical and thermal parameters of the new axial-flux disc-motor in conformity with Fig. 1., with respect the well known radial flux bell-rotor DC motor.

	Axial-flux	Radial-flux
Nominal voltage	12	12 V
Resistance of coils	15,8	16 Ω
Max. power	2,27	2,25 W
Max. efficiency	72	80 %
Speed constant	816	674 min^{-1}V^{-1}
Torque constant	1,17	1,4 mNcm/mA
Stiffness	11,0	7,61 min^{-1}mNcm^{-1}
Inertia	3,4 gcm^2	1,0 gcm^2
Thermal resistance	18 K°/W	33 K°/W
Thermal constant of the rotor	180 s	7 s

The most electrical and electromechanical features of the new developed motors are the same as the conventional cylindrical ones, except the mechanical inertia, and the thermal constant of the rotor. The mechanical inertia is approximately 4 times greater, and the thermal constant is 25 times greater, than at the conventional designed motors.

3. THE NEW METHOD FOR THE INVESTIGATION OF THE VOLTAGES, GENERATING IN THE WORKING ROTOR-COILS

It was not easy to measure the various kind of voltages generating in the rotating coils. For investigation of the commutation-process it was necessary to develop a "reverse sided" motor-model, where the shaft of the rotor is fixed, and the stator with the brushes are in turning.

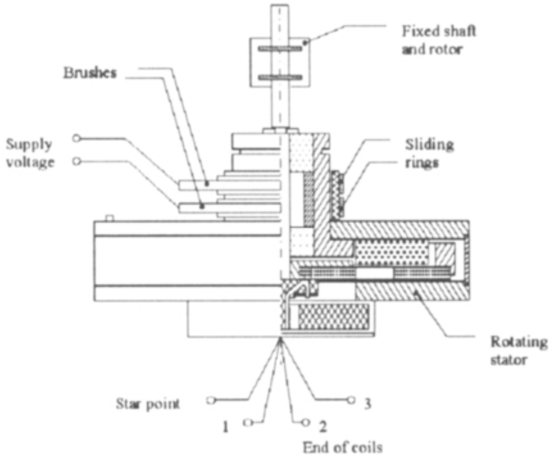


Fig. 2. The "reverse sided" motor model

The current conduction to the rotating stator is solved by two sliding-ring. This "reverse sided" motor works at the same way, as the normal motor works, but the coils and the commutator are standing, and the permanent magnets are rotating. So the ends of coils are simply contacted to standing pins, and by oscilloscope we can observe the voltage of any coil or star-point during operation of the motor. This motor-model also may to load with a breaking-system, and observed the voltages at variable current values. There is a further advantage of this method: if the motor works in a generator-mode, the induced electromotive force in a single coil is observable very well, and from this voltage may to determine the torque-constant of the motor.

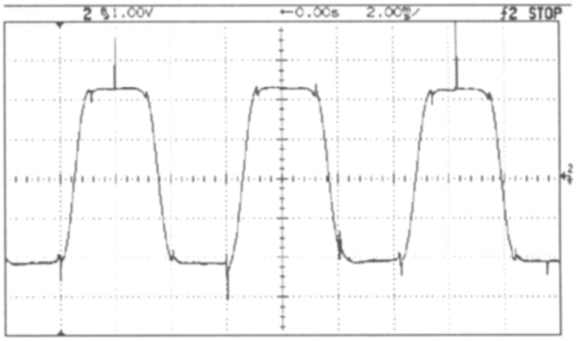


Fig. 3. The induced voltage waveform in one coil

The schematic illustration of "reverse sided" motor-model shows the Fig. 2., and the observed voltage of a single coil is shown on the Fig. 3.

It is visible, that during one revolution the induced voltage two times changes the sign. In this case the motor isn't excited, it works in generator mode.

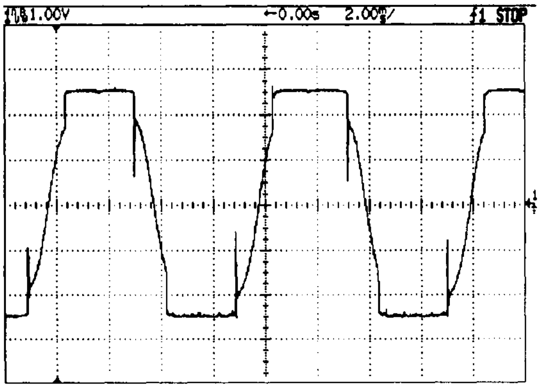


Fig. 4. The voltage waveform in a working coil

If the motor is working, the coil is supplied by exciting voltage four times during one revolution: it gets two times the zero, and two times the supply voltage. (The coil is alternately connect to the supply and to the ground by the brushes.) If the coils isn't excited, the induced electromotive force changes the sign, as shown in the next oscillogram (Fig. 4.).

When the coil is free of current, changes the polarity of the magnetic field. The Figure 4. shows very well the most problematic phase of the commutation process: the interruption of the current. In this case in the coils generates induced voltage, which is depending on the inductance of coil, and the speed of current change. This induced voltage causes the building of sparks between the brush and the commutator segment, and the sparks occur the erosion of both connecting elements. (This process is very similar to the spark machining technology.) Hence, the most important problem is, how can we reduce the induced voltage at interruption of the current. There are two methods for reducing the induced voltage: the application of damping capacitors, or varistors. We have investigated the question of the damping capacitors. The most simply solution of the question is, if we make with 3 capacitors and 3 coils a second-order system. In this case the capacitors are working as energy-storage elements, so reducing the induced voltage. On the next figures show the influence of the capacitors. The Fig. 5. shows an oscillogram, recording with an storage HP oscilloscope at rotating "reverse sided" motor. On the figure is well observable the instant of the interrupt, and the shape the induced voltage, when the capacity was only 1,5 nF. The next figure (Fig. 6.) shows the same oscillogram, but the capacity here 100 nF was. One may be seen, that the shape of the induced voltage are in the both figure very similar, but the first maximum was 15 Volts in Fig. 5., and only 1,75 Volts in the Fig. 6.

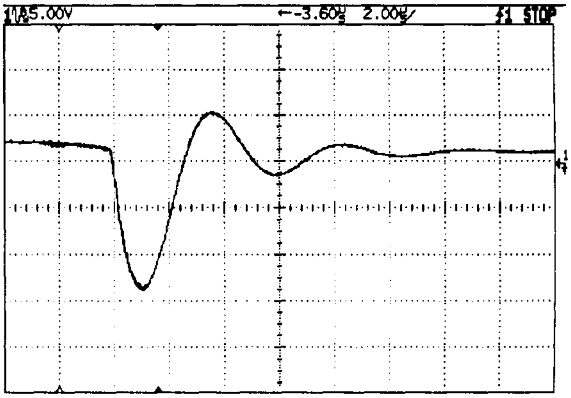


Fig.5. The voltage waveform by 1,5 nF damping capacitor

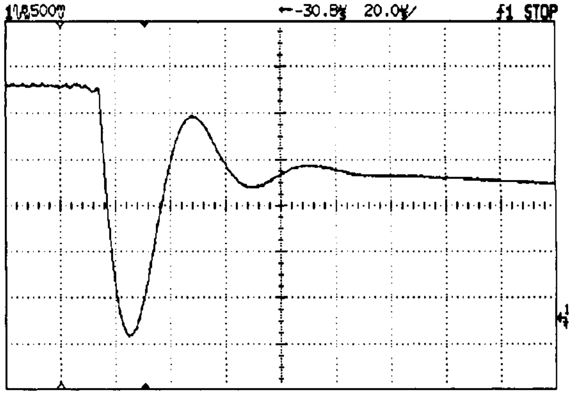


Fig. 6. The voltage waveform by 100 nF SMD capacitor

Consequently, the first maximum of the induced voltage is reduced very well by the damping capacitors, which are integrated on the coils. With the aid of this method may be reduce the arcs, and so to increase the life time of the motor.

The further increasing of the damping capacitor has a limit: during the commutation process, if the brush switches on the coil with the shunt capacitor, it generates a short, but very large current peak, because in the first second the capacitor reacts as a short circuit. For the protecting this unwanted current peaks, may be apply series resistances, which are in the series connection with the damping capacitors. These resistors limit the charger current of the capacitors. The Fig. 7. shows such a diagram.

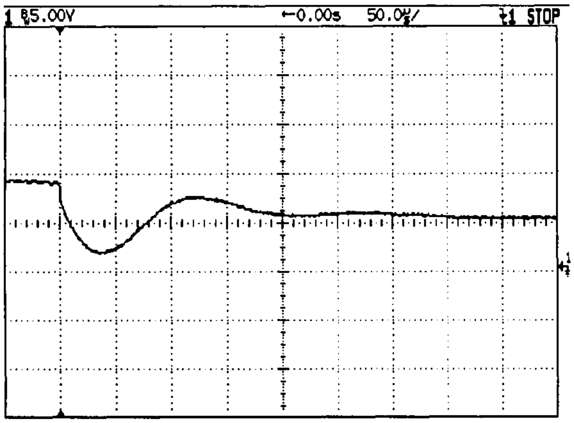


Fig. 7. The voltage waveform by 220 nF SMD capacitor and 22 Ω series SMD resistor.

It is observable, that the damping factor of the second-order system has increased by the series resistor.

4. LIFE-TIME AND RELIABILITY TESTS

The traditional DC servo motors has a generally problem: the commutator brush contact and wear of the both elements. Therefore we have made some experiments for testing only the commutator-wear. The time of these tests was 100 hours in all case, but the load was different. Namely:

- only mechanical connection without electrical load,
- poor ohmic load with a non-inductive resistance,
- normally inductive load,
- normally inductive load with damping capacitors,
- the same case, but lubricated with special grease.

The results are presented on the next slides. It can be seen, that the electroerosion effect is the greatest in the third case, and the damping capacitors effectively reduce the electrical arcs, and the wear of the commutator and brushes.

5. CONCLUSIONS

Based on calculations and measures, it is observable, that the mechanical and electromechanical parameters of the new and traditional motors are similar, but the thermal constant of the new rotor has multiplied respectively to the well known motors. This difference have been resulted from the construction to the coils integrated ring-formed heat-sink, and that means some advantages to the applications: namely the new motor has a greater resistance against overloading, and consequently have a higher reliability.

The new disc motor design developed by the author combined with a similar disc-formed harmonic-drive, or a planet-drive offers a new possibility to drive of various kinds of robots or other mechanisms.

REFERENCES

1. Janocha, H: Aktoren. Grundlagen und Anwendungen. Springer-Verlag Berlin, Heidelberg, New York 1992.
2. Richter: Elektrische Stellantriebe kleiner Leistung. VDE Verlag GmbH, Berlin, Offenbach 1988.
3. Rajki: Törpe és automatikai villamosgépek. Műszaki Kiadó, Budapest, 1990
4. Moczala: Törpe villamos motorok és alkalmazásaik. Műszaki Kiadó, Budapest, 1984
5. L. Molnár, Dr. A. Halmai: Mechanical Connection Analysis of a Micromotor Commutator-Brush. Proceedings of International Conference on Recent Advantages in Mechatronics, ICRAM 95, August 14-16, 1995 Istanbul, Türkey.
6. Dr. A. Halmai: Új konstrukciójú elektromágneses aktuátorok a mechatronikában: a tárcsás forgórészű egyenáramú törpemotorok. Habilitációs értekezés, Bp. 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Fast Actuation of a Hydraulic Switching Valve by Parametrically Excited Structures

M. Garstenauer and Prof. Dr. R. Scheidl

Institute of Mechanics and Machine Design, Johannes Kepler University of Linz, Austria
e-mail: michael@mechatronik.uni-linz.ac.at, scheidl@mechatronik.uni-linz.ac.at

The mathematical model and several design ideas for a new type of hydraulic switching valve are presented. Design goals are extremely high switching frequencies (in the order of 1 kHz), and robust and simple mechanical components and control electronics. The actual flow control element is driven by a parametrically excited elastic structure, in this case a beam. A method that allows an appropriate design of this beam is developed. A first prototype with hydraulic force control was built and its set-up is presented. Several design ideas for future, high speed versions are discussed focusing on piezoceramic actuators, the elastic structure, and the flow control element as key components.

1. INTRODUCTION

The introduction of energy efficient switching techniques (as discussed [1-3]) which are already commonly applied for industrial electric drives to hydraulic drives currently fails due to the lack of a high speed, robust design, and low cost switching valve. There are several research projects dealing with the development of high speed hydraulic valves (see e.g. [4]).

A common problem of these approaches is the fact that the forces which are necessary to accelerate and decelerate the motion of the flow control element rise quadratically with increasing operating frequencies. To achieve fast operation one has to bring high amounts of kinetic energy in and out of the system during a short period of switching time. This fact yields huge and expensive switching actuators.

2. OPERATING PRINCIPLE

One way to overcome this problem is to provide a means of energy storage within the system where the necessary kinetic energy for the transitional motion can be stored during quiescent phases. Elastic displacement of an appropriate structure is one way to achieve this energy storage effect.

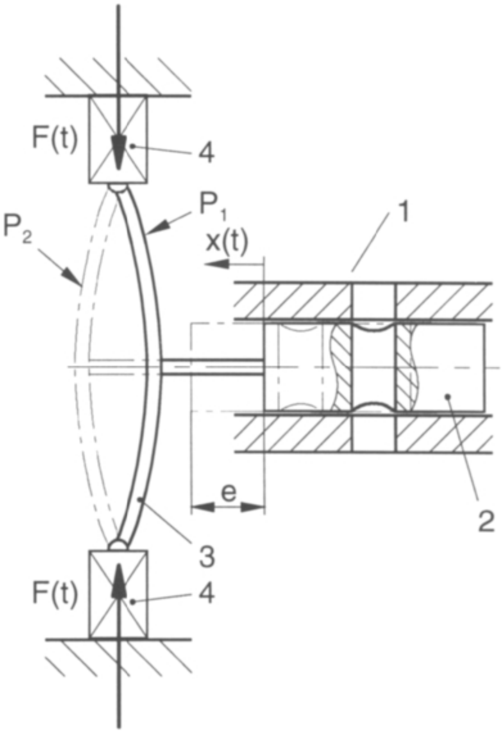


Figure 1. Structure of the switching valve.

Figure 1 shows the structure of the switching valve under discussion. The actual valve body (1) houses a flow control element (2) which is attached to an elastic beam (3). Two actuators (4) exert an axial force on this beam thus keeping it in one of its buckled equilibrium positions P_1 where the valve is open.

To initiate a transition to the closed position P_2 the force F is released and thus the beam and the attached flow control element start to oscillate predominantly in the first eigenmode of the system. When the desired position P_2 is reached (i.e. after half an oscillation period), a force value F is applied again to keep the control element in the P_2 position for any desired time. The same procedure can be applied for the reversal switching operation.

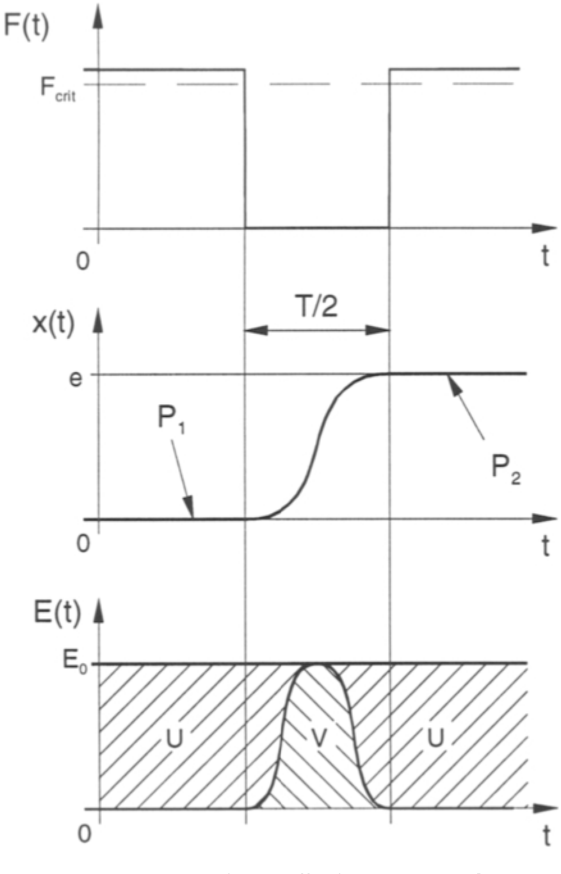


Figure 2. Idealised force, displacement, and energy functions during one transition.

Another important feature is the transformation of force/displacement characteristics from high displacement/low force to low displacement/high force paving the way for direct piezoelectric actuation.

Figure 2 shows the course of actuation force F(t), displacement x(t) and total system energy E(t) for the idealised case (no dissipative forces, infinitely fast actuation).

One can see that after releasing the actuation force to zero the actuator begins a sinusoidal motion towards the other dead centre. The reversal motion is halted by again loading the beam with a force that is slightly higher than its critical buckling load F_{crit}. The energy diagram shows the main advantage of this design. During the switching process the potential energy U that is stored within the beam as elastic displacement energy is used to accelerate the moving elements. After reaching the point in the middle of the two dead centres the kinetic energy V is again transformed to the elastic potential.

In the ideal case this switching process can be achieved without any additional external energy, one needs just to turn on and off the exciting force. Of course, in the real process energy is dissipated due to friction and flow forces, but at high switching frequencies the necessary acceleration forces nominate all others.

. BEAM DESIGN

To achieve a smooth transition between holding and motion phase the beam has to be shaped appropriately. For structures showing identical buckling and vibration mode shapes this requirement is fulfilled. This section discusses a general method to determine the beam shape that achieves equal mode shapes.

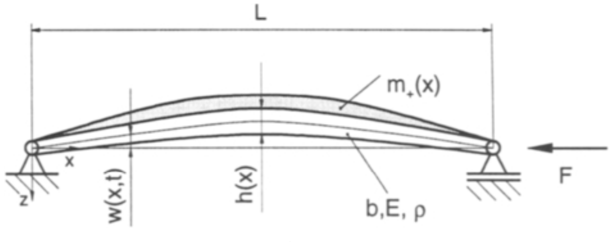


Figure 3. Buckling and vibrating beam.

3.1. Formal Problem Statement

Given are a simply supported beam with length L, an added mass distribution along the beam axis $m_+(x)$ (with no contribution to beam stiffness) and known material parameters (Young's Modulus E, density ρ) as shown in Figure 3. For practical reasons we consider beams with a rectangular cross section with constant width b along the beam axis.

We seek the beam height function h(x) for which buckling mode shape and vibration mode shape are equal. Furthermore, we want to know this shape function φ(x), the buckling force F_{crit} and the first natural frequency ω_I of the beam.

3.2 Fundamental Beam Equations

Both beam buckling and beam vibration are classical problems of elastostatics. Fundamental relations describing them can be found e.g. in [5].

The static buckling of a beam can be described by the following linear fourth order ordinary differential equation.

$$\left(E I w''\right)'' + F w'' = 0 \tag{1}$$

Beam vibration is governed by a linear fourth order partial differential equation.

$$\left(E I w''\right)'' + \left(\rho A + m_+\right)\ddot{w} = 0 \tag{2}$$

Both problems show the same boundary conditions (the argument t refers to the vibration problem only)

$$\begin{array}{ll} w(0,t)=0 & w''(0,t)=0 \\ w(L,t)=0 & w''(L,t)=0 \end{array} \tag{3}$$

We represent the added mass as a multiple of the beam mass by the mass function Ψ(x).

$$m_+(x) = (\Psi(x) - 1)\rho A \tag{4}$$

3.3 Analytical solution for beam height h(x)

In the case of idealised conditions stated above the beam motion is harmonic in time, so one can substitute a separation of variables ansatz

$$w(x,t) = \varphi(x)\sin(\omega t) \tag{5}$$

into the equation of motion (2) resulting in the following equation for the shape function φ(x).

$$\left(E I \varphi''\right)'' - \omega^2 \Psi \rho A \varphi = 0 \tag{6}$$

(1) (using φ instead of w as displacement variable) and (6) show a similar structure. Subtracting (6) from (1) yields

$$\varphi'' + \frac{\omega^2 \Psi \rho A}{F}\varphi = 0 \tag{7}$$

The buckling equation (1) can be integrated twice considering the boundary conditions (3) resulting in

$$\varphi'' + \frac{F}{EI}\varphi = 0 \tag{8}$$

(7) and (8) now form a system of ordinary differential equations with equal boundary conditions. The shape function φ has to solve both equations simultaneously. This means that both coefficient functions have to be equal yielding the algebraic relation

$$h(x) = \sqrt[4]{\frac{12F}{\omega^2 \Psi b^2 \rho E}} \tag{9}$$

for the desired height function h(x). Substituting (9) into (7) or (8) (both yields the same result) gives the following Sturm-Liouville Eigenvalue Problem

$$\varphi'' + \omega\sqrt{\frac{12\rho\Psi}{E h^2}}\,\varphi = 0 \tag{10}$$

Solutions for (10) exist if certain conditions on the coefficient function hold which are fulfilled in this case (see e.g. [6]).

3.4 Numerical solution of the eigenvalue problem

To be able to solve the eigenvalue problem (10) one first needs to determine the mass function Ψ(x) from the given mass distribution $m_+(x)$. This is achieved by combining (4) and (9) to form a fourth order polynomial which can be solved numerically. Uniqueness of the solution follows from feasibility conditions on Ψ, i.e. Ψ real, Ψ ≥ 1.

The actual solution of the Sturm-Liouville problem is performed numerically. According to [7] a finite difference approach was implemented to discretise (10) at N+1 equidistant points in the interval [0,L], yielding a generalised matrix eigenvalue problem

$$\mathbf{A}\,\varphi + \omega \mathbf{B}\,\varphi = 0 \qquad (10)$$

with **A** being a tridiagonal matrix and **B** a diagonal matrix of the coefficient function of (10), evaluated at each discretisation point. It can be solved easily using commerical numerical software packages[1] yielding the natural frequency ω and the respective shape function φ.

3.5 Results

To verify the presented method, a benchmark problem was analysed both using the above method and a commercial finite element package[2]. Table 1 shows the parameters of the problem.

Table 1
Parameters of the benchmark problem.

Beam	Load mass
$L = 0.25$ m	located at the beam centre
$b = 30$ mm	$m_0 = 50$ g
$h_0 = 4$ mm	diameter 7.2 mm

An analysis for a beam with constant thickness h_0 was carried out as reference problem with the FEM package (Case 1). Then an appropriate beam thickness function was used and the first natural frequency ω_1, the critical load Fcrit and the shape function were determined using both the numerical solution method developed above (Case 2a) as well as an FEM analysis (Case 2b). Results for these three cases are given in Table 2.

Table 2
Results for the three test cases.

Case	ω_1 [s^{-1}]	F_{crit} [N]	max ($\lvert\varphi_B - \varphi_E\rvert$) [1]
1	787,5	5306	$6.16 \cdot 10^{-3}$
2a	659,5	3698	-
2b	658,7	3705	$1.12 \cdot 10^{-6}$

One can see a very close match both in the results for the first natural frequency and the critical buckling load. The relative difference in mode shape can be reduced by more than three orders of magnitude by the presented design method.

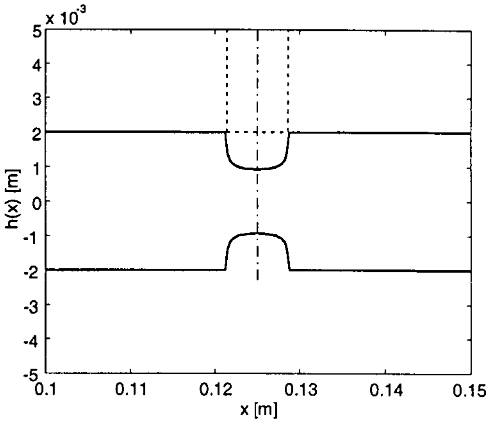


Figure 4. Height function h(x) at the beam centre.

According to (9) the beam thickness is constant wherever there is no added mass. That means in our case the beam has only a small notch in its centre, i.e. where the valve piston is attached. Fig. 4 shows a detailed view of the centre of the beam whose thickness is determined according the method presented above. The dashed lines show where the added mass is attached.

The corresponding mode shape is shown in Figure 5.

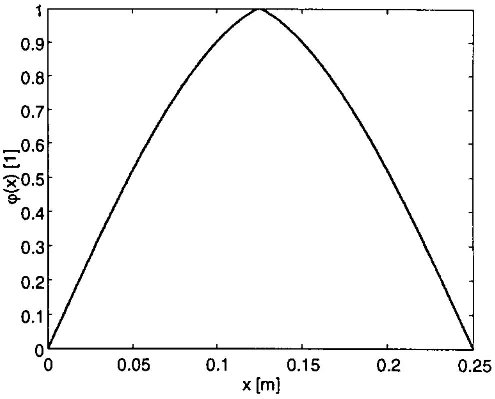


Figure 5. Mode shape φ(x) for test case 2.

---

[1] Matlab™ 5.1 for Windows NT

[2] ABAQUS™ 5.7 for SGI

4 FIRST PROTOTYPE

To verify the feasibility of a valve with this operating principle a first prototype was designed and built. Figure 6 shows a block diagram of the prototype.

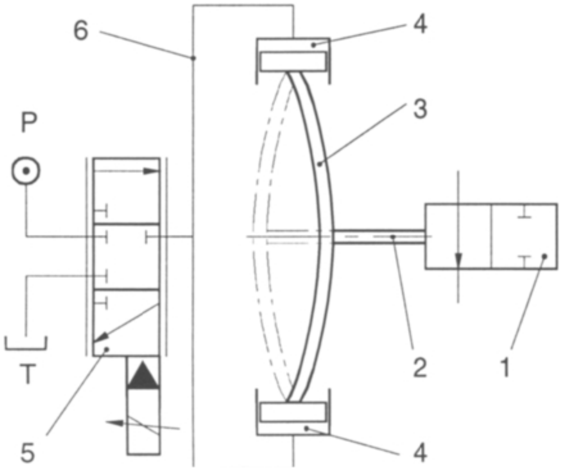


Figure 6. Block diagram of the first prototype.

A conventional size 6 switching valve (1) was adapted to be able to externally connect the actuating beam to the valve piston by attaching a rod (2) to the piston. The beam (3) is loaded axially by two symmetrically mounted hydraulic pistons (4). The load is controlled by a high speed hydraulic servo valve (5). Great care was taken that both pressure chambers and the supply lines (6) from the valve to both pistons are symmetric. This is vital to achieve a synchronous load distribution on both pistons. There are no hinges at the joint of the piston and the beam as the sealing element can deflect in a sufficient range of displacement ensuring the simply supported boundary condition.

The key design parameters were chosen according to the following data.

Table 1
Design parameters of first prototype.

Total mass of the flow control element m_0	50 g
Switching displacement e	3 mm
Beam length L	0.2 m
Maximum/minimum beam height h	3 / 1.3 mm
Beam width b	30 mm
First natural frequency of the system f_1	100 Hz

Due to the limited dynamics of the hydraulic force control unit the system's natural frequency was chosen to be around 100 Hz.

6 HIGH SPEED DESIGN

In this section several ideas for a follow-up prototype are presented. An overview of the design is given in Figure 7.

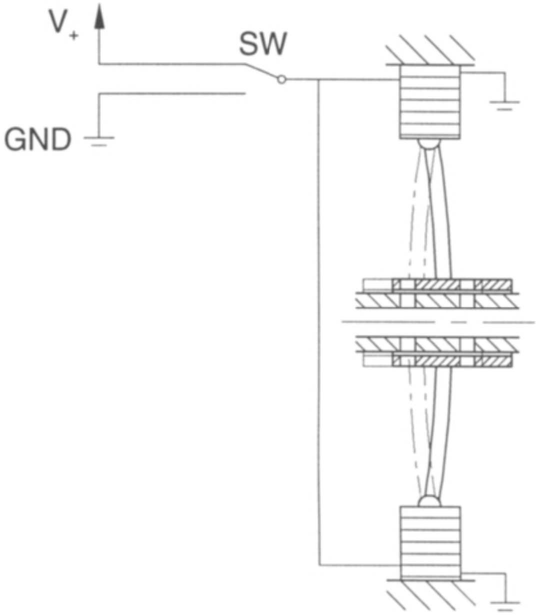


Figure 7. Design proposal for a high speed version.

6.1 Actuation

Conventional hydraulic piloting stages are not capable of achieving switching frequencies in the order of 500 Hz to 1 kHz. Piezoelectric stack elements will be used instead. The piezo-actuator is directly connected to the elastic beam and thus oscillates with the same natural frequency. Using the switch SW this oscillation can be stopped by applying a voltage across the piezo element. Only the power that is necessary to overcome the dissipative losses due to friction and fluid forces has to be delivered by the electric power supply.

6.2 Flow Control

The conventional switching valve (valve body and flow control element) is replaced by a jacket/tube type design. It features low necessary displacements and low mass of moving parts which are both key features for high switching frequencies. A detailed view of this design is given in Figure 8.

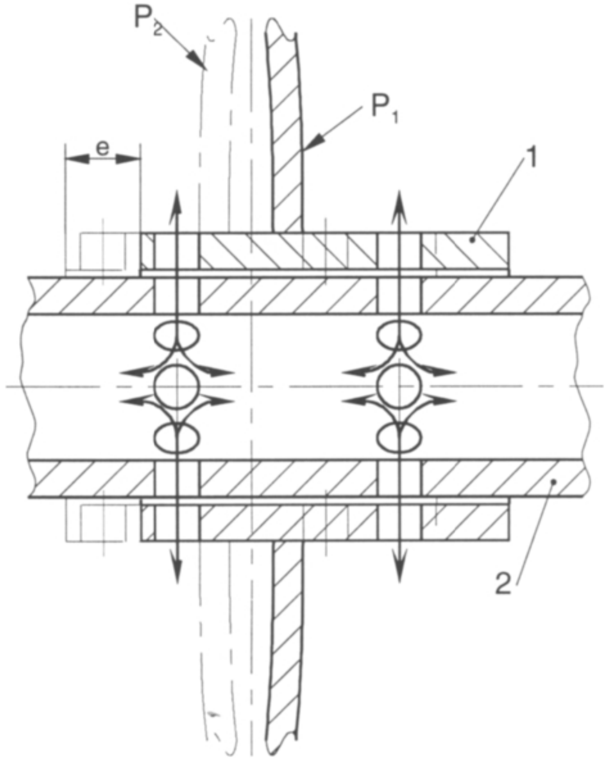


Figure 8. Detailed view of the jacket/tube type flow control element design.

A jacket (1) is directly attached to the elastic beam and is able to slide on a tube (2). Both the jacket and the tube feature several holes that coincide in one switching position (the open position shown in Figure 7) and overlap their counterparts in the other, i.e. closed, position. Bores with a diameter of down to 0.5 mm are feasible to produce, and thus the switching displacement e can be reduced to around the same value. The mass of the moving jacket will be in the order of 10 to 20 grams. A challenging design issue is the joint between the elastic beam and the jacket. It should link both parts as rigidly as possible in the switching direction but on the other hand not hamper beam deformations. Otherwise important assumptions used in the beam design process are not met.

7. SUMMARY & OUTLOOK

A new type of hydraulic switching valve was presented. It takes advantage of a parametrically excited elastic structure (in this case a beam) both as energy storage and transformer of force/displacement characteristics. First a general design algorithm to determine the shape of the elastic beam was developed. A first prototype was designed and built to demonstrate the feasibility of the proposed principle. Several design ideas for a next generation prototype were discussed. Key components are high speed/high energy efficient actuation and a lightweight flow control element with a small necessary switching displacement.

REFERENCES

1. H. Gall and K. Senn. Freilaufventile - Anwendungskonzept zur Energieeinsparung bei hydraulischen Linearantrieben, Ölhydraulik und Pneumatik, 38/1994, Nr. 1-2.
2. R. Scheidl, D. Schindler, G. Riha, and W. Leitner. Basics for the Energy-Efficient Control of Hydraulic Drives by Switching Techniques, Proc. 3rd Int. Conf. on Mechatronics and Robotics, Paderborn, Oct. 4-6, 1995.
3. R. Scheidl, M. Garstenauer, and S. Grammer. The Resonance Converter - A Novel Method for Hydraulic Fluid Power Control, Proc. 5th UK Mechatronics Forum International Conference, Guimarães, Sept. 18-20, 1996.
4. R. Kasper, J. Schröder, A. Wagner. Schnellschaltende Hydraulikventile mit piezoelektrischem Stellantrieb, Ölhydraulik und Pneumatik, 41/1997, Nr. 9.
5. F. Ziegler. Mechanics of Solids and Fluids, Springer Verlag, New York - Vienna 1991.
6. H. Heuser: Gewöhnliche Differentialgleichungen: Einführung in Lehre und Gebrauch. B.G. Teubner, Stuttgart 1995.
7. J. Pryce: Numerical Solution of Sturm-Liouville Problems. Oxford Science - Claredon Press 1993.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

The piston-in-cylinder electromagnetic ram

Eur. Ing. Phillip Denne B.Sc. C. Eng. C.Phys. F.I.E.E. M.Inst.P M.R.I.
Guilden Ltd. Bournemouth, England.

The paper describes the leisure-industry origins of a new and useful form of linear motor, outlines some of its unusual properties and points to a number of applications in manufacturing industry and civil engineering

SIMULATORS

Ask any non-technical person what a simulator is and the answer will be about teaching pilots to fly aircraft. Mention entertainment simulators and you will learn that the most important thing about them is that they move. To the general public a simulator is not a simulator unless it moves - and a machine that rocks about a bit in an amusement arcade is "a simulator".

But I am talking about properly-engineered simulators - machines whose purpose is to present a totally-convincing artificial reality. Of course, entertainment simulators do not actually simulate anything real. They are fantasy machines in which the laws of nature are modified so that the experience that the public pays to receive does not closely resemble any physical reality. Instead, the machine portrays reality as the common man would prefer it to be - at least for three exciting minutes.

ELECTRONIC WARFARE

Twenty years ago I designed and built a very large, complex and expensive simulator - and a special building to house it - for Marconi at Stanmore. This was designed to test various types of electronic warfare equipment. Electronic Warfare is a game of Hide and Seek, a sort of lightning-fast blindfold mobile three-dimensional chess game played between computers and microwave equipment at extremely high speed and without the real-time intervention of any human being. For example, radar systems and jamming equipment will try to outwit each other as an aircraft comes streaking in across enemy territory whilst the defending radar-controlled equipment tries to get a fix on the aircraft and fire off guns or missiles to knock it out of the sky. The jammer must win the EW battle if the aircraft is to survive. The electronic warfare simulator is designed to test to the limits in a real-time scenario the efficiency of EW equipment hardware and software.

The problem is that everything happens very quickly, is complex, silent and invisible. It was difficult for me to explain to VIP visitors what was happening - or rather what was being simulated in real time. So I decided that I would try to transform the microwave signal parameters into visible images which could be shown on a TV screen. In effect I decided to present a virtual reality view of the microwave world as it would be "seen", for example, by the microwave sensors in an attacking aircraft.

I was astonished to discover how exciting this visual simulation was, for me and for the visiting VIPs and senior military personnel from many countries who watched the displays on the TV screens. Clearly, those normally dour and serious individuals were fascinated and found themselves caught up in what they saw. I resolved that I would find some way to allow the general public to experience that sort of excitement. So I left Marconi and set up a company which built its first "Super X" Venturer simulator in 1985.

ENTERTAINMENT SIMULATORS

Super X "Venturer" machine

A typical entertainment simulator consists of a small cinema in which a group of between twelve and forty people watch a projected television display and listen to a soundtrack, whilst the whole cinema moves on a hydraulic motion base. As I discovered later, the idea was first conceived by Douglas Trumbull, the special effects film producer, in 1978.

If you have watched a simulator working from the outside you will see that it weaves and dodges about through small angles of pitch and roll and it also bobs up and down a few inches, apparently in a random way. But I assure you that when you get inside the capsule the effect is quite extraordinary. In a closed environment, it is impossible to tell the difference between gravitational and inertial forces - in short, you do not know which way up you are and, since your body ignores a steady acceleration after the first fraction of a second, it can be deceived by a transient which is then "wiped out". Thus it is possible for a "motion simulation choreographer" to persuade the occupant of a capsule which only moves a few inches that he/she is, for example, swooping violently around the sky, racing around the Isle of Man on a motorbike, skiing in the Alps, rushing through a canyon on a raft, or sailing gracefully about in a microlight aircraft.

The essence of simulation - of Virtual Reality - is that it should be a compelling fantasy. That is, during the progress of the simulation it must appear to be "real", even if you understand that it wasn't real *when you think about it afterwards*. If you have never ridden a good simulator then I can assure you that you will be very surprised just how compelling the experience can be. Thorough simulation is capable of disconnecting you thoroughly from the real world and immersing your mind within a Virtual World, in which the laws of physics can be different and in which you need not be limited to being the self that you are in your life outside the simulator. Entertainment simulators are very profitable!

THE PROBLEMS OF HYDRAULICS

The growth of the entertainment simulator business soon created a strong demand for simulators that could be placed in indoor locations, in small rooms in upper storeys of well-furnished buildings, close to the general public. But the fine oil mist that always emerges from hydraulic machinery quickly causes severe damage to nearby fabrics, constitutes a fire hazard and has been alleged to have toxic effects. Further, hydraulic motion systems can impose very high shock loads on a building structure and they produce a lot of heat. Insurance companies charge very high premiums for such machines and it is expected that the US health and safety authorities will soon ban their use in certain locations. I set out to find a better way to make a motion base.

GAS SPRINGS

I decided to replace the hydraulic rams by gas springs. The gas springs support the deadload and they act as reservoirs in which energy may be stored during the first part of each movement and from which it can be extracted during the return stroke. The actuator rams now only need to provide impulsive forces. There is no requirement for a steady upthrust. Power is only consumed when the simulator is in motion and the only power required is that necessary to change the kinetic energy of motion and to overcome any frictional losses.

ELECTROMAGNETIC RAMS

Since the forces must be coupled to a floating mass the actuators *have to be electromagnetic*. There is no other way to take a firm grip on a freely-suspended inertial object. Because no suitable electromagnetic rams were available to us in 1991, we had to develop a new type of actuator, which now turns out to have many other applications.

What we have done, in concept, is to slice a rotary electric motor down to the middle, roll it out flat and then roll it back up again by taking the long edges of the strip and bringing them round to form a cylinder. We do the same with the armature and we put the armature shaft down through the centre of the cylinder to make a piston. What we now have looks just like a pneumatic or hydraulic ram but is actually an electric motor.

Lotus-designed ram in 1994

INHERENT FORCE-MEASUREMENT

A few years ago I visited Simultronix in Los Angeles and for the first time I rode a motorcycle simulator that was fitted with pressure pads as part of a servosystem that allowed the rider to throw the machine about by changing his body position whilst being subjected to the appropriate visual, haptic and audible motion cues. It was very impressive.

Some of the most exciting and enjoyable things that we do have a strong element of body movement control in them - and they are some of the most difficult and dangerous to learn. They are large markets for training simulation that involve body motion. For example:-

Skiing, ski-boarding, ski jumping, skidoo racing
Ice skating, roller skating, skateboarding.
Surfing, wind surfing, sail-boarding, jet ski racing, water skiing.
Sailing, yachting, white-water canoeing.
Tobogganing, Bobsleigh racing, Motorcycle racing, Horse riding, Hang gliding

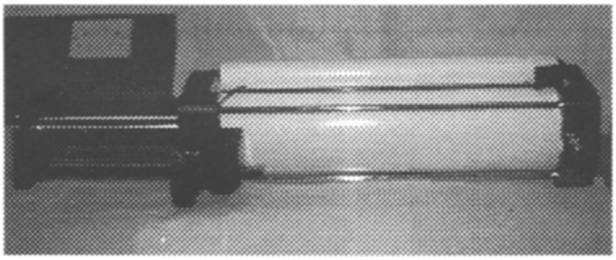
When a human occupant of a simulator on an electromagnetic motion base moves or leans to left or right, fore or aft, the currents flowing in the rams will immediately adjust themselves to the degree necessary to compensate for the shift in the centre of mass, holding the position of the motion base constant. Thus by monitoring the relative values of the currents flowing in the rams, it is possible to know the position of the c.o.g. of the moving platform - that is, to sense any movements of the human occupant(s). The movements can then be fed back into the control computer to modify the progress of the simulation.

To illustrate the concept of body motion control we arranged for a local champion surfer to pose with a VR helmet on a motion base, and we later proved the principle in a complete working machine that was demonstrated on "Tomorrow's World".

The "Cybersurfer" concept

INDUSTRIAL APPLICATIONS

The publicity that resulted from the entertainment demonstrations of the ram soon led to a surprising variety of enquiries from engineers working in other industries. The ram was therefore redesigned mechanically and its electromagnetic systems were reconfigured to meet the demands of industry. It is now generally called a "ServoRam", since it is electrically equivalent to a three-phase AC servomotor. It is possible to construct rams of this design that have peak thrusts from 10 Newtons to at least 200 KiloNewtons, each of which is compatible with a range of standard industrial electric motor drive units.



Three-phase AC ServoRam

The industrial rams have several advantages:-

No stray fields. The "piston-in-cylinder" topology means that all the fields are contained within the ram, so that there are no spurious effects on its surroundings and no problems with swarf or steel grit.

Cleanliness. Modern versions of the ram are fully sealed. They will operate under water, in a vacuum or in a sterile atmosphere.

Silence. Most other types of electric linear actuator are noisy and the noise level increases with thrust rating and with wear. This can be a problem in machinery that operates continuously and in close proximity to a human being. It is especially distracting in a clean, high-technology environment.

Simplicity and reliability. Nothing moves except the output element. Nothing wears except two bearings and one air seal.

Wide operating tolerances. There is no requirement for micron filtration or white-glove maintenance.

High peak-to-mean thrust ratio. There is no hard limit to the transient peak thrust capability of the ram over a short time interval.

Speed of response. This is an order of magnitude better than the best hydraulic device.

Power regeneration. Using the ram in reverse mode, the machine operates as a generator - a controllable damper with an electrical power output that may be returned to the supply system.

Ease of control. Thrust is strictly a linear function of drive current, which may be made the controlled parameter and which is independent of the position of the piston. There is zero time delay between current flow and output thrust.

Precision. Zero backlash, zero transport lag and a wide control bandwidth lead to extreme precision, which is limited only by that of the position transducer. A precision better than one micron can be achieved.

Dual Action. Very significant amounts of power can be saved by using the pneumatic function of the ram to handle the static or quasi-static loads and the electromagnetic function for precise positioning.

Auto-tuned spring function. Similarly, it is possible to combine a gas-spring technology with the electromagnetic forces to save power and to speed up precision reciprocating machinery.

Scaling. The topology is scaleable over a wide range of thrusts and stroke lengths, using off-the-shelf components.

EXAMPLES OF INDUSTRIAL APPLICATIONS

Passenger Lifts. Many passenger lifts in buildings up to five storeys high are raised by hydraulic rams. Electromagnetic rams can replace them with an efficient, silent, clean and much more reliable product. Changing demographics have created a market for a lift module that may be added to existing homes and it is expected that the price will fall so that lifts can be fitted as standard in general private housing.

Automobile Suspensions. Several makers of luxury cars are considering the use of electromagnetic suspension systems. It is rather too early in its development for the device to meet the requirements of a power element in active suspensions but, acting as a damper, it can vary its characteristics under computer control in milliseconds if required. That is, an electromagnetic suspension system can be instantly tuned for every road condition as it is encountered. What is more the energy absorbed by the damper (sometimes 300 Watts a wheel) is not thrown away as heat but can be fed back into the battery supply, saving fuel. Simultaneously, the fluid piston element can act as a self-levelling, height-adjusting fluid spring.

Rail Coach Suspensions. Mathematical studies carried out by Loughborough University as part of an EEC programme show that the ideal suspension for a rail vehicle is a combined gas and electromagnetic system. Other work in association with two major European companies shows that the necessary forces, displacements and power levels can now be achieved.

Industrial Automation. The advantages of extreme precision, cleanliness, silence, high peak thrust, unlimited speed, intrinsic force-sensing, reliability and wide control bandwidth mean that the device has many factory applications. For example, the rams can replace hydraulics in food and drug manufacture, where contamination is very expensive. In high-speed machinery such as that for sorting, transferring and packing goods, the unique ability of the electromagnetic actuator to "freewheel" when handling inertial loads saves a great deal of energy.

As an enhancement we have recently developed an extension of the motion base auto-tuned gas spring technology which may be used when handling reciprocating mechanisms. This reduces still further the electrical power demands and allows higher operating speeds.

Doors and Vents. Although vent actuators only operate at infrequent intervals and their first cost must be low, they are often placed in almost inaccessible positions, so maintenance is very expensive and reliability is vital. The ServoRam is interesting because it has only one moving part - the piston itself.

Security. The power needed to operate a ServoRam can be stored in a small space - for a year or so in a battery or for several minutes in a charged capacitor. This means that the device can be self-contained, so that it may be used to carry out a security action in a no-power emergency. It would be possible for the device to drive a bullet-proof security screen into position faster than a robber could pull the trigger.

Stabilised Platforms. When optical devices like cameras or measuring equipment have to be mobile on a land or sea surface they need to be isolated from disturbances when they are in use. This is also true for guns, radar antennae and missile launchers. Previously, most stabilising arrangements have used "hard" mechanisms with powerful motors, gears and cranks to generate the movements that are necessary. There is considerable interest in using the ServoRam gas-spring suspension to decouple the stabilised platform from external disturbances and the very fast electromagnetic forces to hold a precise position.

Cross-country vehicle seating. Human beings need stabilisation to avoid fatigue and injury when driving vehicles such as earth-moving equipment across rough terrain for long periods. Although soft spring suspensions have already been tried, it is clear that a significant improvement would result from the use of electromagnetic actuators that combine an air-spring and an active position-stabilising function.

Pile Driving. As a result of earlier studies by a British architectural consultant, it seems to be feasible to use both the fast rise-time and the force-sensing feedback properties of the new ram in a novel form of high-speed pile-driver. This works by applying a powerful impulsive waveform to the pile at a frequency that adapts automatically to the characteristics of the soil and makes it more fluid.

Bulldozing, Planing and Ploughing. In the same way, it seems to be possible to reduce the mechanical forces used in earth-moving equipment of various kinds.

PRACTICAL LIMITATIONS OF THE TECHNOLOGY

Weight. There is no mechanical gearing. Whilst the machines are simple in concept and have only one moving part they have, at this stage of their development, a low energy density. As a result they are heavy, because they need plenty of copper for the current and plenty of steel to conduct the flux around it. But if the application and its environment will allow the ram to be overdriven hard, so that it can run hot (even up to about 200°C), the weight can be reduced significantly.

If we assume
flux density B as 0.5 Tesla
current density d as 5 MegAmps/sq. m
density of copper as 9 tonnes / cu. m
copper resistivity as 1.56 * 10^-8 ohm m

then it is easy to show that for practical, conservatively-rated designs the weight of a typical ram will be approximately 50 Kg per 1000 Newtons continuously-rated static thrust

Efficiency. The power efficiency is speed-dependent, but in complete contrast to fluid rams it *increases* with velocity. At maximum power output (half maximum thrust, half maximum speed) existing rams are about 60% efficient. Rams can be produced for velocities up to at least 5 metres / second with current technology

Peak power demand Using the above assumptions, it is easy to show that the typical power requirement of a conservatively-designed ServoRam is approximately 350 watts per 1000 Newtons continuously-rated static thrust.

For most applications we arrange for the peak static thrust to be at least four times the continuous rating. But since the thrust is directly proportional to drive current, the resistive losses are proportional to the square of the thrust. The peak power demand of a typical ram is therefore 5.6 KW per 1000 Newtons continuously-rated static thrust - and the motor drive unit should be rated for this.

DUAL ACTION

In connection with the foregoing, it should be remembered that the electromagnetic system has been designed to operate with a gas spring or even a conventional pneumatic function. The rams are usually capable of exerting pneumatic forces at least twice as great as their peak electromagnetic forces.

Moving a 1 Tonne mass at 0.5 g over a short distance (say 0.5 metre) only requires a ram that is rated for a peak thrust of 5000 Newtons - perhaps only 1000 Newtons continuous static thrust.

It should also be remembered that the ram is capable of acting remotely from the load and being connected to it by a fluid (e.g. liquid) driven by the ram piston / armature.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

SMART ULTRASONIC SENSOR FOR SEMI-AUTONOMOUS ROBOTS

R.Kažys

Kaunas University of Technology, Ultrasound Research Center
Studentu 50, Kaunas 3031, LITHUANIA

The structure and principle of operation of the smart ultrasonic sensor for semi-autonomous mobile robots and assembly carriers. are presented. The measuring methods and signal processing techniques are discussed.

1. INTRODUCTION

Semi-autonomous intelligent mobile robots offer great potential in hazardous environments in applications such as nuclear plant maintenance / decommissioning, surveillance, mine clearance, and bomb disposal. They must have appropriate sensing systems enabling to detect obstacles, to measure their coordinates and in ideal case to distinguish between different obstacles. In order to meet these requirements the sensor must possess some intelligence, e.g. to be a smart sensor. For these purposes ultrasonic sensors are videly used [1-5].
Ultrasonic sensing systems are rather attractive due to their low cost, but there are some physical limitations that impose restrictions on a performance of such systems. The most fundamental is due to a low value of the ultrasound velocity in air, which reduces the information update rate. That starts to be critical in the case of mobile robots. These problems can be overcome using novel measuring techniques combined with an appropriate digital signal processing.

2. PRINCIPLE OF OPERATION

In order to achieve the above mentioned goals the sensor should possess some intelligence, enabling the vehicle to navigate safely in a dynamic environment. For this purpose the sensor should have ability to adapt to changing conditions, for example, to modify a survey sector , to select particular directions for more detailed analysis of surrounding objects, etc. Most suitable solution is to use an ultrasonic adaptive smart sonar, which collects the information around a mobile vehicle by means of an electronically steered phased array.
The generic block diagram of the sensor, fulfilling the above-mentioned requirements is presented in Fig.1.
In order to survey space in two planes - vertical and horizontal- the array should be two-dimensional. Control of the array and the whole sonar is performed by microprocessors incorporated into the sensor. Communication of the ultrasonic sensor with other subsystems of the robot is performed via LON Works type field bus.
The use of electronically controlled arrays allows to perform all operations in the real time, what enables the sensor to adapt to a dynamic environment. The commands required for the adaptation are obtained from the host computer via the LONWorks interface.
Using conventional techniques survey of surrounding space and determination of the coordinates of obstacles is performed by means of a narrow ultrasonic beam, which is sequentially scanned in the region of the interest (Fig.2). This can be performed either mechanically or electronically , but in the case of 2D survey only the electronic scanning should be used. The main problem is that the time necessary for performing such a scan is relatively long due to a low value of ultrasound speed in air. A single direction measurement requires a time

$$t_s = \frac{2l_{\max}}{c_a} + \Delta\tau \ , \qquad (1)$$

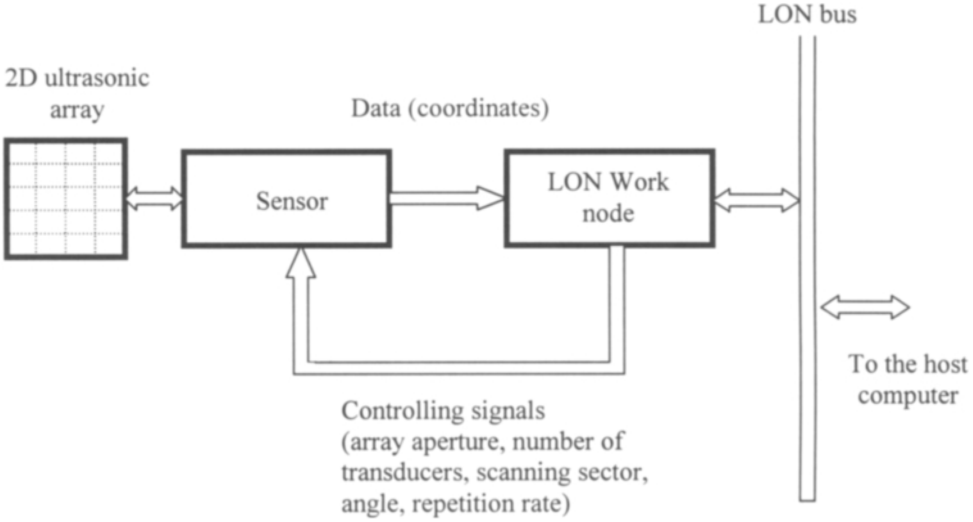


Fig.1. Generic block diagram of the adaptive ultrasonic ensor

where l_{max} is the maximal distance, c is the ultrasound velocity, $\Delta\tau$ is the additional interval between measurements in order to avoid an influence of multiple reflections. It is necessary to point out that a presence the multiple reflections is acute problem, because it can cause a false detection of the objects which actually do not exist. Therefore, usually the interval between two subsequent measurements is increased by the time interval $\Delta\tau$ until the all multiple reflections reduce to a negligible level.

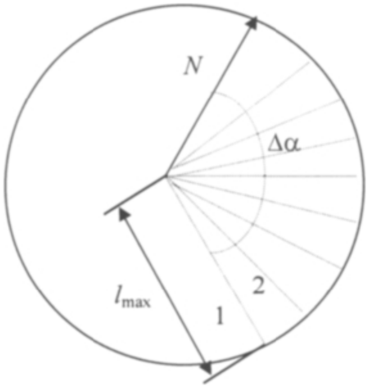


Fig.2. Sequential scanning of environment

The complete scan will take a time

$$T_s = N\left(\frac{2l_{\max}}{c_c} + \Delta\tau\right). \qquad (2)$$

Here N is the number of ultrasound beam positions:

$$N = \frac{\Delta\alpha}{\Delta\beta_{0,5}}, \qquad (3)$$

where $\Delta\alpha$ is the scanning sector and $\Delta\beta_{0,5}$ is the width of the directivity pattern at the 0,5 level.

If $\Delta\alpha$=180°, $\Delta\beta_{0,5}$=10°, N=18, l_{max}=5 m, $\Delta\tau$=10ms, then the duration of the single complete scan of the 0 -180° sector takes a time T_s=0,72 s. During this time a mobile vehicle moving at the speed v_{max}=0.8 m/s will travel l= 0.58 m. It means, that due to a movement of the vehicle and a relatively low rate of measurements, the coordinates of objects detected will be measured not in the absolute coordinate system, but in the coordinate system locked to the moving vehicle.

As it follows from Eq.2, and 3, the speed of measurements can be increased in the following ways:

- reduction of the time required for a single direction measurement t_s;
- reduction of the number of scans N;

The first method is based on a reduction of the additional time interval $\Delta\tau$. This approach is illustrated by Fig.3. In the case of conventional measurements at different directions identical ultrasonic pulses are sent periodically (Fig.3a). The interval between two adjacent pulses $Tr.P_1$ and $Tr.P_2$ is chosen according to Eq.1.

Duration of this interval can be reduced using for different directions different orthogonal coded sequences. The transmitted ultrasonic pulses $Tr.P1$, *Tr.P*2, *Tr.P*3 are orthogonal coded sequences, for which the following conditions are fulfilled:

$$E_{ij}(T)=\int_0^T f_i(t)f_j(t)dt=\begin{cases}E, & \text{if } i=j\\ 0, & \text{if } i\neq j\end{cases} \qquad (4)$$

Here *i, j* are numbers of the transmitted pulses, *E* is the energy of the pulse, $f_i(t)$, $f_j(t)$, are the waveforms. For this purpose such quasirandom signals as the Golay codes or the *M*-sequences can be used. In this case signals transmitted at different directions have different signatures. Due to the orthogonal properties of the signals caused by transmitted pulses at different directions, the multiple reflections received at the direction 2, but caused by the pulse transmitted at the direction 1 can be efficiently suppressed by means of the correlation processing.

At *i*th direction the signals reflected by objects and caused the "correct" *i*th transmitted pulse at the output of the receiver are given by

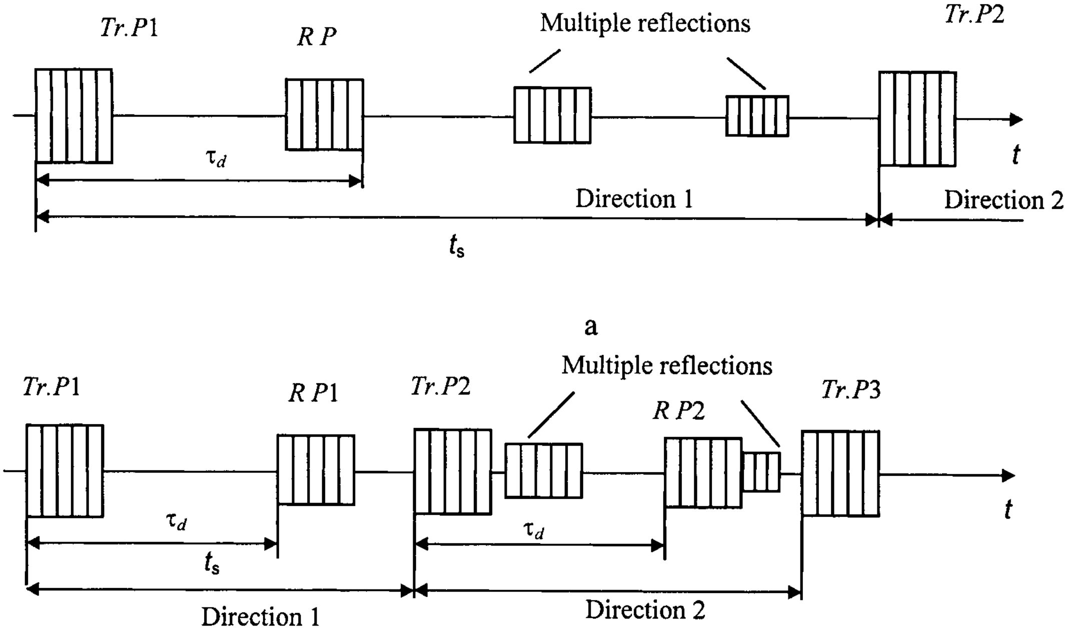
$$u_{out}(t)=\int_{t_1}^{t_2}\left[\sum_j u_j(t)\right]f_i(t)\neq 0. \qquad (5)$$

Correspondingly, at (*i*+1)th direction for the multiple reflections caused by the pulse transmitted at *i*th direction we shall obtain

$$u_{out}(t)=\int_{t_1+t_s}^{t_2+t_s}\left[\sum_j u_j(t)\right]f_{i+1}(t)\approx 0, \qquad (6)$$

e.g., due to orthogonal properties of the signals at the adjacent directions, the signals caused by multiple reflections are suppressed. In such a way the time required for a single scan can be reduced up to a few times depending on the features of the signals used In the sensor developed as such coded signals phase modulated binary M-sequences and Barker codes are used.

The second approach enabling to reduce the time of a complete scan is based on increasing the width of the ultrasonic beam $\Delta\beta_{0,5}$ used for measurements. In order to avoid the loss of accuracy of measurements of the bearing angle, the position of the reflector is found using the tri-aural perception technique (Fig.4). If the object detected is inside the directivity pattern of the ultrasonic transducers, the



a

b

Fig.3. Data acquisition by means of an ultrasonic sensor using the conventional (a) and coded (b) signals at different directions: Tr.P are the transmitted ultrasonic pulses, R.P are the signals reflected from the obstacles

distance and the bearing angle can be found from a single measurement [4, 5]. In order to determine the position of objects in 3D space, two sets of ultrasonic transducers located in two perpendicular planes are required (Fig.4). One plane is parallel to the surface of the floor on which a mobile vehicle is moving, another plane is perpendicular to the first one. The transmitter in Fig.4 is indicated by the gray dot.

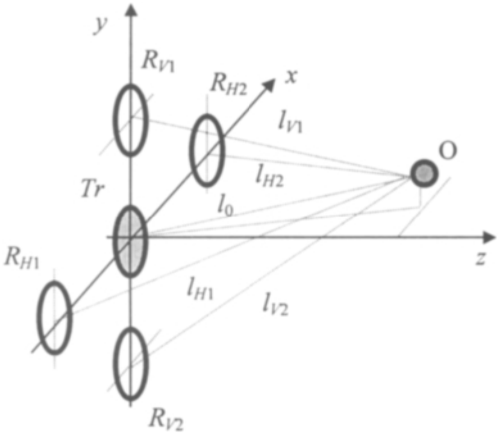


Fig.4. Determination of the position of the object *O* in 3D space using the tri-aural approach: *Tr* is the transmitter and R_{H1},R_{H2}, R_{V1}, R_{V1} are the receivers in horizontal and perpendiculat planes.

The distance l_0 is determined from the measured delay time t_0 , which equals to the time required for an ultrasonic signal to propagate from the ultrasonic transmitter to the reflector and back to the receiver:

$$l_0 = \frac{ct_0}{2} \qquad (7)$$

The bearing angle α in one plane, for example, horizontal one, is given by

$$\alpha = \arcsin\left[\frac{l_{H2}^2 - l_0^2 + 2r_c(l_{H2} - l_0)}{2d(l_1 + r_c)}\right]. \qquad (8)$$

Here r_c is the diameter of the reflector and d is the distance between receivers in the same plane.

The distance to the object and corresponding bearing angles in vertical planes are found in the same way as described above. In order to collect an information in the 0 - 180° sector, it is necessary to rotate electronically the ultrasonic beam radiated by the transmitter and correspondingly to change the directions of reception. In this case the transmitter and each receiver must be not a single ultrasonic transducer, but electronically steered phased ultrasonic arrays.

3. STRUCTURE OF THE SENSOR

The sensor consists of 4 ultrasonic transmitters, 5 ultrasonic receivers and appropriate electronic units for generation of driving coded signals, amplification of the received signals and parallel signal processors (Fig.5). The delay time of the signals reflected from obstacles is determined using correlation signal processing. This processing is carried out in each receiving channel simultaneously by means of TMS320C50 type signal processors. That enables to carry out measurements at single direction during 200 ms. The correlation processing ensures also a good noise robustness of the sensor in a noisy environment [8].

4. BEAMFORMING BY MEANS OF ULTRASONIC PHASED ARRAYS

For transmission of the ultrasonic signals electronically steered phased arrays are used. Each array consists of a few low cost MURATA piezelectric transducers operating at 40 kHz. The number of the transducers in one array can be up to 8 elements. The number of the elements in the array and the driving signals can be selected using or controlling signals generated by the signal processor, or by the host computer. It means, that arrays become adaptive. One array scans the sector 0 –90°. The directivity of the array is chosen in such a way, that it is possible to survey this sector during 3 subsequent scans. That enables to survey 0 –180° sector during 0.5 s, even taking into account the time necessary for the processing. Reception of the signals reflected is performed by 5 individual MURATA type transducers located at different places. That is required for the tri-aural processing.

The calculated directivity patterns of the array possessing 8 elements at 3 different deflection angles (0°, 10° and 20°) are presented in Fig.6a. In a pulse-echo mode the directivity of the whole system is obtained multiplying the directivity pattern of the array by the directivity of the receiver. For this reason a level of the side lobes in pulse-echo mode is significantly lower. It is necessary to point out also, that the non-linear tri-aural processing used for estimation of the coordinates, influences the

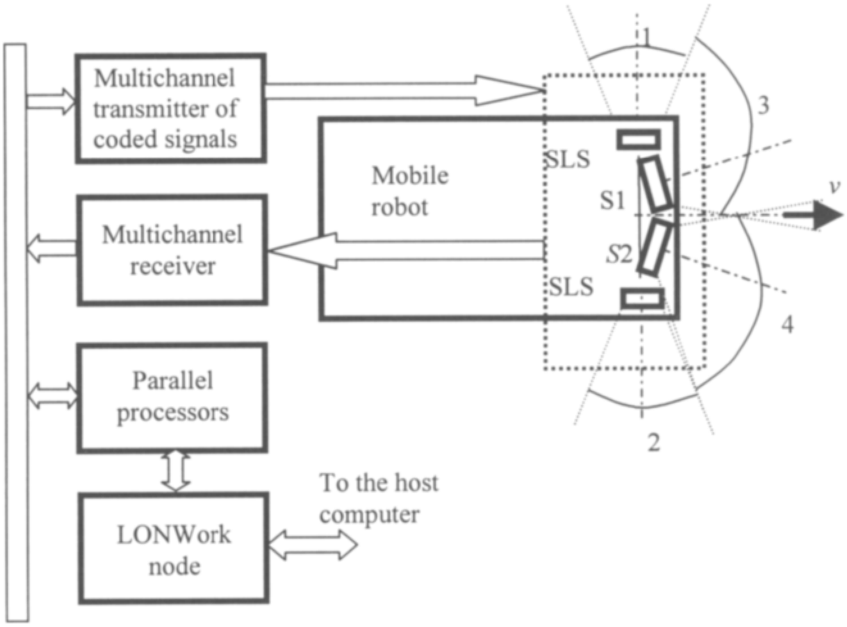

Multichannel transmitter of coded signals
Multichannel receiver
Parallel processors
LONWork node
Mobile robot
To the host computer
SLS
S1
S2
SLS
1
2
3
4
v

Fig.5. The structure of the ultrasonic sensor: SLS are the side looking sonars, S1, S2 are the main electronically steered arrays, 1, 2, 3, 4 are the areas covered by corresponding sonars.

directivity of the whole sensor and reduces the level of the side lobes.

The measured directivity pattern of the 8 element linear array in a transmitting mode at 0° is presented in Fig. 6b. It is in a good correspondence with the pattern calculated. The wider side lobes and a small asymmetry are due to variations of the geometry of the array from the ideal one.

The width of the beam is controlled selecting the number of elements in the array. Direction of the beam is governed electronically by introducing delays of the driving binary coded signals. The necessary delay time is given by:

$$\tau = \left(\frac{d_\Delta}{c}\right)\sin(\beta), \qquad (9)$$

where d_Δ is the distance between adjacent elements, β is the deflection angle.

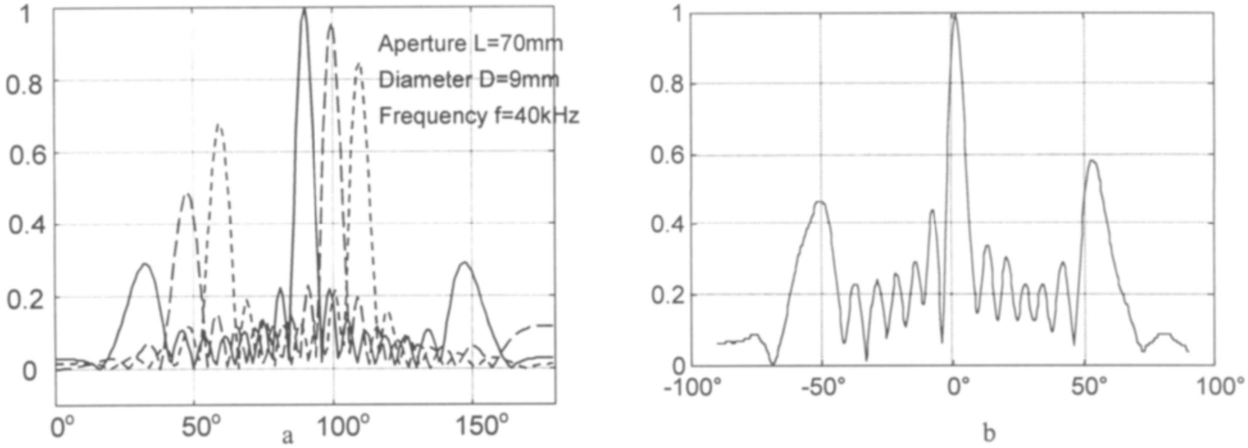


Fig. 6. Resulting directivity patterns of the transmitting 8 element array: a –calculated, b-experiment.

EXPERIMENTAL RESULTS

The performanc of the sensor has been tested using variuos reflectors. The examples of the signals obtained from the cyllindric type reflector are shown in Fig.7.

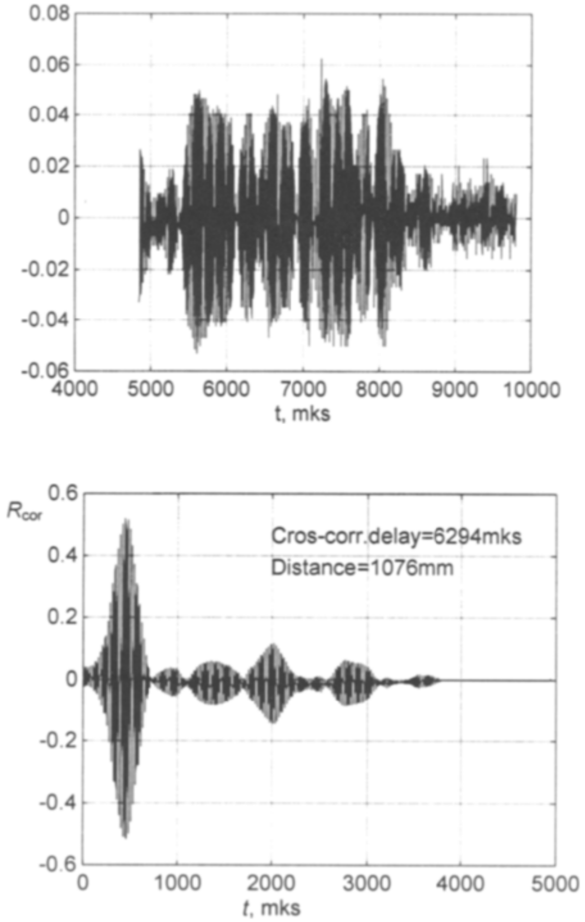


Fig.7. The raw and processed signals (phase modulated 11 element Barker code) reflected by a 10mm diameter rod at 1076 mm from the sensor

5. CONCLUSIONS

The ultrasonic sensor proposed is based on the following main principles:

- the sensor is the multichannel instrument, which detects obstacles and measures their coordinates simultanously at a few diffrent directions. That increase the speed of operation and reduces dynamic errors;
- surrounding objects are detected using coded wide band ultrasonic pulses and correlation processing;
- location (the coordinates) of the objects are determined using tri-aural approach together with an electronic scanning of the ultrasonic beam in space;

Use of electronically controlled arrays allows to perform all operations on-line, what enables the sensor to adapt to a dynamic environment. The commands required for the adaptation are obtained from the host computer via the LONWorks interface.

6. ACKNOWLEDGMENTS

This project was carried out as a part of INCO-COPERNICUS project No.ERB IC15-CT96-0726.

REFERENCES

1. M. Drumheller. Mobile robot localization using sonar. IEEE Trans. on Pattern Anal. and Machine Intelligence. Vol. 9, No.2, 1987, p.325-332.
2. R. Magori. Ultrasonic presence sensors with wide range and high local resolution. IEEE Trans. on Ultrasonics, Ferroelecterics and Frequency Control, No.2, 1987, p.202-211.
3. R. Kažys, K.Kundrotas, V. Dzimidavicius, L.Mažeika, A. Borkowski. Programmable ultrasonic range finder for mobile robot. Robotersysteme., Vol.7, 1991, p.101 -106.
4. R.Kuc. Biomimetic sonar recognizes objects using binaural information. J. Acoust. Soc. America, vol. 102, No.2, Pt.1, 1997, p. 689 - 696.
5. H. Peremans, K. Adenauert, J.M. Van Campenhout. A high resolution sensor based on tri-aural perception. IEEE Trans. on Robotics and Automation, Vol.9, No.1, 1993, p.36-48.
6. P.Webb, I.Gibson, C.Wykes. Robot guidance using ultrasonic arrays. Journal of Robotic Systems, Vol.11, No.8, 1994, p.681 -692.
7. R. Kažys. Delay time estimation using the Hilbert transform. Matavimai (Measurements), Kaunas, Technologija, Vol.3, No.1-2, 1996, p.42-46.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Mechatronic Development Environment for Precision Engineering

Prof. Dr.-Ing. Dr. h.c. Hans Grabowski, Dr.-Ing. Stefan Rude,
Dipl.-Ing. Zsolt Pocsai, Dipl.-Ing. Mei Huang

Institute for Applied Computer Science in Mechanical Engineering (RPK)
University of Karlsruhe; Kaiserstrasse 12, 76131 Karlsruhe, Germany

Abstract: The crucial points in the design of multi-disciplinary products will be discussed, with emphasis on the use of heterogeneous software tools. To overcome these problems, a three layer integration, i.e. the process-, the model- and the system technical integration, is considered as the practical solution. The integration on the model level will be described more in detail. Further on, such an integrated mechatronic development environment which supports the product development in the area of precision engineering will be presented.

Keywords: design theory, computer aided design, mechatronics, product model, model development and integration, precision engineering.

1. Introduction

The implementation of design methodology is not sufficient to provide a development environment for mechatronic products. Two main reasons are:

A. Design objects are changing.
The products concerned in classical design methodology are usually one-discipline-artifacts, e.g. pure mechanical artifacts, while the multi-disciplinary products, namely mechatronical products, are becoming popular. In other words, the classical design approaches do not deal with the multi-disciplinary engineering knowledge which is essential nowadays for market competition.

B. Design tools are improving.
With the development of computer technology, more and more software is used in daily design work. From text editor to engineering database, from 2D/3D CAD to finite element analyzing, they not only improve the work efficiencies but also change the way engineers work.

While academia therefore recognises the need of further investigations the direction of a design theory, industry is using guidelines for product design in different branches. For example, the Society of German Engineers (VDI) has defined the VDI-guideline 2221 as mechanical product design methodology and VDI-guideline 2422 for the design of instruments with microelectronical control function.

This kind of guidelines only cover one special area of the industry until now. Therefore, a systematic design methodology for the mechatronic products as an industry guideline is queried. One of the goals of the project „EntwicklungsUmgebungen MECHatronik", (EUMECH, stands for „Development Environments for Mechatronics", sponsored by the German Federal Ministry of Education, Science, Research and Technology (BMBF)) is to prepare the necessary information and knowledge for a future guideline of mechatronic product development which is intended to be a VDI guideline. A number of research institutes and industry partners such as BOSCH, BMW and Siemens-Nixdorf participate in EUMECH. The authors are also involved in this project.

In this paper we focus on how current information technology will be used for the design of multi disciplinary products, i.e. what a mechatronic design environment should look like. An integrated development environment for mechatronic products in the area of precision instruments will also be presented.

2. Characteristic attributes of a mechatronic design environment

According to the main working areas of mechatronic shown in Fig. 1, the research of mechatronic design environments belongs to the process-oriented research, which is also the main topic of design methodology. Design methodologies and environments which systematise the product development process have been recognized in the past few years as a key technology for product development using different engineering domain knowledge. These methodologies include schedules which connect work steps or development stages, rules and strategies for a resource-efficient realization as well as development-supporting methods and tools. Currently there is a lack of an integrated development methodology which might support the development of complex multi-disciplinary (by means of mechatronics) products.

When designing complex multi-disciplinary products it is naturally assumed that the existing supporting tools are various and heterogeneous. The historical reasons of such kind of "ununity" are just the same as for computer technology: different hardware and software vendors, different internal (company-specific) and common standards, protection of know-how and copyright, etc. Such a situation keeps existing and should exist if an active and competitive growth of information technology is required. On the other hand, keeping different small supporting tools only for one or two special problems is more practical also from the financial aspect: the users would rather prefer to choose exactly those tools which they need and combine them together than to buy a monolithic tool in which some functions will never be used.

Integration is the best solution to cope with heterogeneous tools. For the mechatronic development environment three integration layers can be considered: the process-, the model- and the system technical integration. The process integration concentrates mainly on how to effectively control the design steps (processes) inside a company or incorporated companies while the model integration tries to couple the different software (tools) on a common semantic level. The key point of system technical integration is to guarantee the information exchange with a series of system services such as CORBA, JAVA or OLE.

The result of integration on the process level is the formal description of the mechatronical design theory in which the process model consists of the multi disciplinary design steps and development stages. The outcome of the model integration is a mechatronical model (see Fig. 2), which links branch or application specific models together on a semantic level. Through the realization of the underlying integration platform, the focus will be set on the integration of the domain specific models towards an integrated mechatronical product model.

Product-oriented research	**Process-oriented research**
• Finding new modules • Finding improved product structures • Design of innovative products • Determining the boundary conditions for practical implementation	• Extending the design theory for mechatronic products - Design process - Module-finding methodologies • Building new development support tools - Product models - Integration platforms

Figure 1. Main research areas within mechatronics

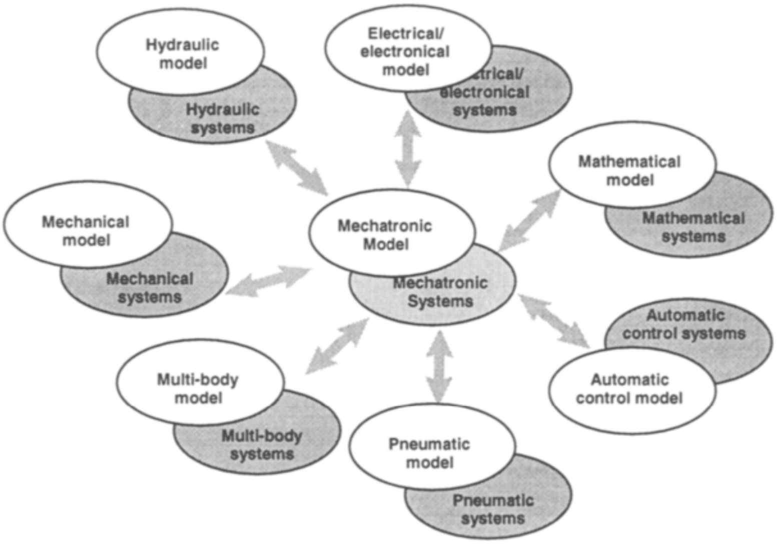

Hydraulic model
Hydraulic systems
Electrical/ electronical model
Electrical/ electronical systems
Mathematical model
Mathematical systems
Mechanical model
Mechanical systems
Mechatronic Model
Mechatronic Systems
Automatic control systems
Automatic control model
Multi-body model
Multi-body systems
Pneumatic model
Pneumatic systems

Figure 2. Domain specific models and systems vs. a mechatronic model and system

3. Common model based mechatronic development environment

Basic requirements and functionality of a mechatronic development environment will be explained with the example of a real industry design task: a needle printer will be designed. It shall work at a lower noise level and with a frequency spectrum which is more comfortable to the human ear. In order to identify all relevant data input the analysis and understanding of the process chain is necessary. This input respectively its content shall be considered in the data model as well at a later time.

The corresponding process chain for the development and parameter optimization is shown in Fig. 3. The geometry is designed with the use of a commercial CAD-system. In order to get acoustic parameter, modal analyse will take place both in laboratory and by software; the computational and the experimental result will be compared and the corrected mode values will be delivered to the acoustic simulation phase. The acoustic behavior of the printer will be calculated by simulation software as well as determined by experiments in laboratory, and subsequently be evaluated. Finally, a modal design will be performed according to the acoustic behaviour - the undesirable noise is tried to be avoided by changing the mode parameter which might lead to changes in geometry (3D model in CAD software).

Behavioural simulation in this context denominates the depiction of a system in a computer model and the operation of such a model in the sense of a virtual prototype. The intended goal is to enable the employed CAE software to directly retrieve the data necessary for its purposes from the integrated data model without the need to start an explicit model for a specific user program.

The simulation includes a sufficiently exact imitation of a technical system's behaviour or parts of it by means of a mathematical model on a computer aiming for the analysis and synthesis of this system. Within the context of the simulation of mechatronic systems, in which models can be observed up to the entire vehicle, not all operations in the involved components/subsystems can be calculated by analytical means because of increased computation effort, on the other side, an analytical model does not always exists. Thus, a *simplified substitute model* is used for systems which can not able to be solved analytically within mechatronic models. Particular attention must be paid to the fact that formalisms to adapt model descriptions towards a mathematical model are very diverse and so far require several different model descriptions. It is exactly this model description which shall be simplified. In the future it will then be possible to use differing formalisms, e.g. for simulation which are all based on a common model description.

For every main function shown in Fig. 3, one or more types of support software are available on the market. But how all of this software practices seamless co-operation, is still a problem: the data interface from one software to another is not clearly defined. On the other side, the data model which is the result of the first software does not cover the full need of the next tool, at least does not supply the same reference to the next tool. Reasons for such a situation: every software has its own data model. This means that the process chain requires a multiple change of system environments, e.g. from CAD system back to modal analysis, resulting in a heavy additional load on the data transfer. Data transfer itself is very time-consuming and often has to be managed "manually". Since the system environment has to be changed several times in the course of a concrete task, high cost and time-consuming additional work arise in consequence.

The common model based integration is the most efficient integration technique in this case. The EUMECH project considers this topic to be an own work package, the MechaSTEP project focuses even more mainly on the task of establishing a STEP-based mechatronic model (MechaSTEP stands for „STEP Data Models for Simulating Mechatronic Systems“ and is also a BMBF founded project with numerous participants from industry and research institutes) The characteristic requirements, which have to be met by the integrated mechatronical models are the following from the user's point of view:

1) *The semantic elements which are used in the domain-specific model should also be included in the mechatronic common mode, i.e:*

Mechanics: solids (rigid and elastic bodies), coordination systems, view point, presentation of geometrical elements.

Electric and electronics: circuit logic, passive components (resistor, capacitor, inductance), active components (current/voltage source, diodes and transistor), circuits connectivity of geometry.

Hydraulics and pneumatics: pumps, volume capacity (compensatory reservoir), transfer elements (valve, tube, etc.), point of view, nodes, tank, graphical presentation of elements etc.

Automatic control: sensors, actuators, control strategy and algorithms.

2) *A common mathematical data model must be established.*

In this mathematical model all domain specific physical quantities should be considered and a unitary

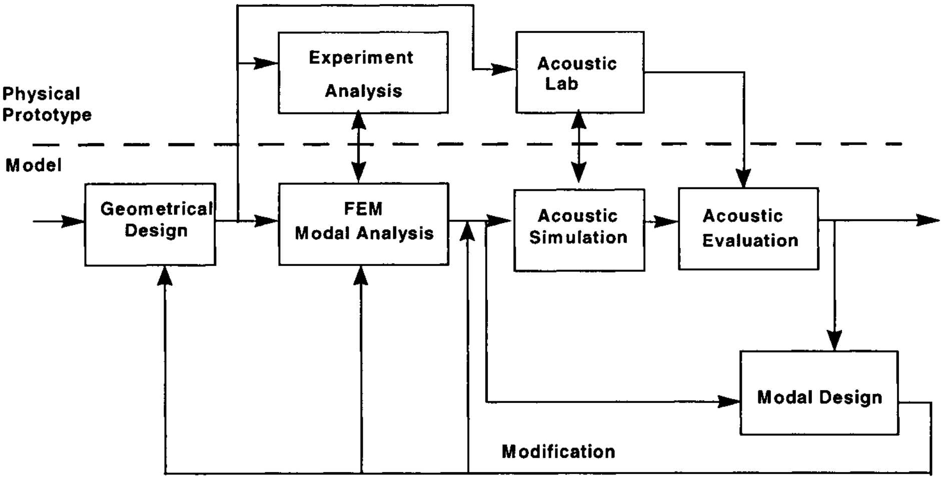


Figure 3. Geometry design and acoustic analysis environment

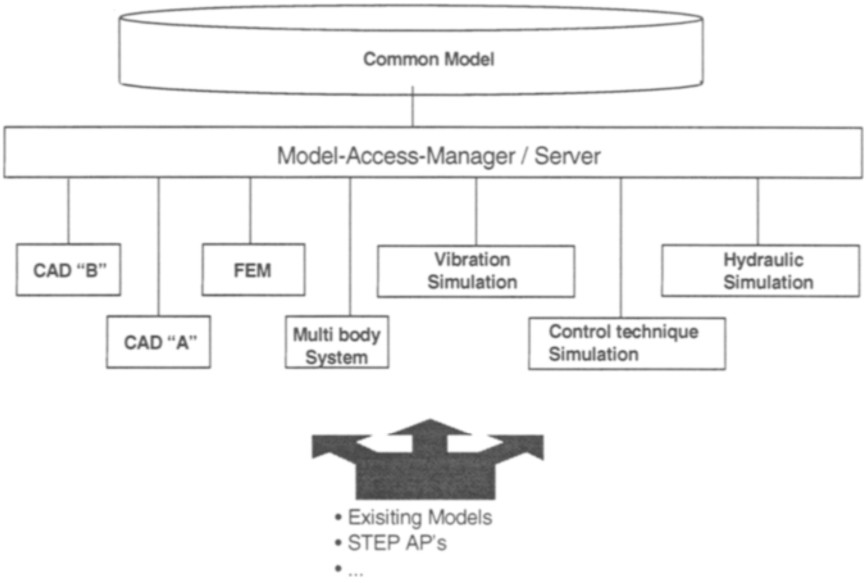


Figure 4. Common model based integration

mapping to the physic dimension must be made. This is the key point for a deeper software integration. The input/ output interface of different software, especially for simulation tools, can be organized as predefined subclass, also the necessary transfer calculation (pre-/post-processor) from one software output to an other's input can be regarded as operations in this mathematical model.

3. Global data should be generalized.

Not only environment temperature, system time, etc. are considered as global data in the integrated mechatronical model; the domain specific functions of all linked domains as well as their interoperability should also be represented. Furthermore, information about the platform on which each software/hardware tool works, must be covered. The platform information includes operation system, required memory, and most important: the compatibility to another platform, which is necessary for the design of an integration development environment.

Based on such a common model in which the above requirements are fulfilled, an integrated mechatronic development model can be built. The system architecture is shown in Fig. 4.

4. Summary

The elaboration of the mechatronic development environment considers not only the merging of design methodologies in different engineering domains but also completing the existing models to a required common model for representing product attributes and its whole life cycle activities. Essential points are: the design of an integration mechanism for IT-software tools; the use of existing mechatronical modules by means of efficient access to latent mechatronical knowledge; the provision of a development environment and finally the practical evaluation of the development environment. In conjunction with the introduction of a generally accepted model for the depiction of mechatronic-relevant data, the essential foundation for the widespread inter-formalism use of model data of mechatronic systems shall be laid. Furthermore, the exchange of data within a company's various areas of development can be simplified. Already, it is currently necessary to often change the different formalisms during the development process. For this purpose, a new model still has to be assembled for each individual formalism. This effort is rendered unnecessary if a standardized data format for integrated mechatronic models is defined. The development process can then be paralleled better and development time thus be shortened. The application of different program packages as well as the conglomeration of partial models into an overall model will not be problematical. Great benefits lie in store behind these developments for the system supplier as well as for the end user.

Acknowledgements

The authors would like to thank all the EUMECH project members for making this paper possible.

References

1. Harashima, F.; Tomizuka, M.; Fukuda, T.: Mechatronics - What is it, Why and How? In: IEEE/ASME trans. Mechatronics, vol. 1, No. 1 March 1996
2. Grabowski, H.; Rude, S.; Langlotz, G.: Aufbau von Wissensspeichern für die Lösungsfindung im Produktentwicklungsprozeß; In: Wandel im Maschinenbau durch Feinwerktechnik und Mikrosystemtechnik - 41. Internationales Wissenschaftliches Kolloqium. pp. 772-778; 23.-26.09.1996 Ilmenau, Germany
3. Gausemeier, J.; Brexel, D.; Frank, T.; Humpert,A.: Integrated Product Development - A New Approach to the Computer Aided Development in the Early Design Stages; In: Proceedings of the Third Conference on Mechatronics and Robotics - From design methods to industrial applications. pp.10-29; Oct. 4-6, 1995 Paderborn Germany
4. Roddeck, W.: Einführung in die Mechatronik; ISBN 3-519-06357-3; 1997
5. Kallenbach, E.; Birli, O.; Saffert, E.; Schäffel, Chr.: Zur Gestaltung integrierter mechatronischer Produkte; In: VDI Berichte 1315-Mechatronik im Maschinen- und Fahrzeugbau; pp. 1 -14; May 11-12, Moers, Germany
6. Schiehlen, W.; Dürr, R.; Neerpasch, U.; Witte, L.: Mechatronik und STEP; In: Produktdaten Journal Nr. 2/1995 pp.12-19
7. EUMECH Technical Annex, 1998, available in RPK, University Karlsruhe, Germany

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronic System Design Methodology - Initial Principles Based on Case Studies

M. Valášek

Department of Mechanics, Faculty of Mechanical Engineering, Czech Technical University in Prague
Karlovo nam. 13, 121 35 Praha 2, Czech Republic, E-mail: valasek at fsik.cvut.cz

This paper deals with the generalization of experience with the development of three different mechatronic systems (machines, products) into a draft of mechatronic system design methodology of similar systems. All three products could be characterized as "controlled mechanical systems". They are anti-sway crane control, semi-active truck suspension, redundant parallel robot.

1. INTRODUCTION

This paper is based on the generalization of experience with the development of three different mechatronic systems (machines, products). All three products could be characterized as "controlled mechanical systems". They are anti-sway crane control, semi-active truck suspension, redundant parallel robot. The similarity of problems, ways of finding solutions and strength of achieved solutions have been analysed and resulted into understanding of generalized steps of design process which could be treated as a draft of mechatronic system design methodology of similar systems.

The generalized steps of design process are described in the second and third sections together with brief description of judgment from case studies to generalized conclusions. The fourth section deals with the brief overview about the case studies of mechatronic products.

2. THREE DIFFERENT VIEWS OF DEVELOPMENT PROCESS

There are three different views from which we can describe the unique development process of the mechatronic system:

- **life cycle of the product** - the design stages of product specification, concept creation, detailed design, production, usage and maintenance (despite the knowledge about the life cycle laws many traditional product developments do not treat all steps of life cycle in details, these are especially the initial steps of market analysis, product specification and product concept generation, these steps are crucial for mechatronic products);
  Case study experience: In all case studies the essential initial start of the development was the finding of an idea that (i) some property or function of the existing product is missing, (ii) some electronic control could create this property or function and that (iii) this property or function would have significant importance on the market of the product. This finding could be a result of preliminary initial iteration: market analysis - product specification - concept creation, however, in all case studies the proper mechatronic product development has started only after this initial idea. Before that to start the product development was very difficult, i.e. the problem specification as "there is the product, improve it by mechatronic approach" is extremely hard.
- **conceptual solution** - the abstract techniques and procedures for design problem solving. There does not exist generally proved complete theory, although one such approach is TIPS (Theory of Inventive Problem Solving) [3, 4]. (the importance of consideration and generation of many product concepts is essential; however in many cases the criteria and requirements conflict occures which is the basis for creative

and inventive solutions and which leads to the importance of conceptual solution procedure based on the studying of Ideal Solution and Conflict Resolution instead of conflict compromise which solves the given problem despite different constraints);

Case study experience: The product to be mechatronic is being always considered as to be a small invention. In all case studies the essential approach for development of mechatronic product which involves the synergy effect was the looking for Ideal Solution and onflict Resolution. This has prevented in all case studies to stop the development process on compromise solution instead going further and really finding significantly improved solution. In all case studies the Ideal Solution was the directly the source of motivation directly applied for solution construction. Probably the idea of Ideal Solution and Conflict Resolution despite all problems discussed in [3, 4] is extremely important for mechatronic products of the class of controlled mechanical systems.

multidisciplinary system development synthesis - modelling and simulation tools are treated as essential techniques for complex and multidisciplinary system development [1, 2] (the modelling is the essential tool for overcoming many interactions within multidisciplinary system design).

Case study experience: In all case studies the role of modelling and simulation for the product design was very important. The amount of interactions of properties within product design and the lack of empirical experience with the new kind of product which is usually being developed was always overcomed by modelling and simulation.

3. KEY STEPS OF MECHATRONIC DESIGN METHODOLOGY

The experience from the development of case studies can be summarized into a sequence of key steps from the above described points of views of development process which have been essential for successful solution of case studies and which represent the essential difference from the traditional product developments.

I. Analysis of the market → detection of missing product, missing or unsatisfactory property or function of existing machine, creating of initial concept of electronic control which could create the missing feature and importance of this feature on the product market.

II. Analysis of abstract structure of possible solutions (conceptual, functional, structural) by different concept generation, inventive engineering, value analysis etc. → different product concepts, essential feasibility of solution concept.

III. Application of problem solving principle: looking for Ideal Final Solution, Conflict Resolution → new concepts, ideally reachable solution, ideal system control concept which could actually create the basis of real control system.

IV. Real Technological Solution, implementation of proposed control system by economically available technology → practical control concept, feasibility of the product.

V. Detail system development → modelling, simulation, parameter synthesis, optimisation, investigation of dynamic interactions.

VI. Solution implementation and materialization → prototype, function verification.

VII. Traditional product development in mechanical, electrical, control, software engineering → product.

These steps look very naturaly and traditionally, nevertheless they represent a real difference from the traditional product development.

- The necessity of I. step is not so high for traditional products. They have their traditional market. The new properties/functions are required for new market for new products. Mechatronic products are new ones.
- The II. step is specifically required for mechatronic products because the amount of possible conceptual solutions is significantly larger than for traditional products. It is caused by many structures enabled by control and software solutions.
- The III. step is according to the accumulated experience the key approach in the described methodology. It could be applied for traditional

products as well, however the control enables the direct application of ideal solutions.

- The IV. step is required by the fact that the mechatronic product integrates different technologies the combination of which must be adjusted. The technology of traditional product is usually fixed.
- The V. step is now applied also for traditional products but the amount of modelling and simulation is substantially higher for mechatronic products.
- The VI. step is also required for traditional products but the function verification of the mechatronic synergy on the prototype is essential.
- The step VII. is necessary for the integration of applied technologies. The new specific technology for mechatronic products does not yet exist.

4. CASE STUDIES OF MECHATRONICAL SYSTEMS DEVELOPMENT

On the practical cases of mechatronic product developments the particular key steps of the draft mechatronic design methodology can be demonstrated. In fact exactly these development steps were generalized into the proposed methodology.

4.1 Anti-Sway Crane Control [5]

I. Removing of residual swinging of crane load can increase the productivity of crane manipulation by almost 100% (e.g. container cranes, saving of manipulation worker).

II. There are possible closed loop control solution with relatively difficult expensive sensor of load position or open loop control solution without load position measurement. The application of inventive engineering proposed the surprising solution of swinging damper on the crane load. New load position sensors based on load motion sensors (acceleration, pulley rotation) have been proposed.

III. The objective of the solution is not to decrease the load swinging because the load swinging is the basis of load motion. The objective is to remove the residual swinging after the manipulation and to control the swinging. If the equations of motion are linearized and thus decoupled the following dynamic equation for one cartesian direction can be obtained

$$\ddot{x} = \ddot{\varphi}_x + \frac{g}{l}\varphi_x$$

where φ_x is the difference between the position of crane carriage x and crane load on the cable length l, g is gravity. There are position and variable controls. In case of velocity control it is required to move the load from the initial stable position

$$x(0) = x_0, \dot{x}(0) = v_0, \varphi_x(0) = 0, \dot{\varphi}_x(0) = 0$$

into final stable position

$$x(T) = any, \dot{x}(T) = v_0 + \Delta v, \varphi_x(T) = 0, \dot{\varphi}_x(T) = 0$$

with the change of velocity by Δv during time interval T. Such trajectories φ_x which satisfy the above boundary conditions are being searched. Suitable functions for such trajectories are trigonometric ones because of following implementation. One such trajectory is

$$\varphi_x = K(\sin(\frac{\pi t}{T}) - \frac{1}{3}\sin(\frac{3\pi t}{T}))$$

Based on known cable length l, required change of velocity Δv in time interval T and maximum allowed acceleration the unknown gain K in this function can be computed. This solution has been compared with theoretical time optimal control with good results. The advantage over it is that the control is a smooth one. Therefore it is an ideal smooth time suboptimal control.

IV. The derived control from previous step is an example of generalized input shaping control [6] for open loop control. The trigonometric functions can be implemented as a reentry software enabling continuous change of desired velocity [7]. This open loop control trajectory can be considered as the programme trajectory for the closed loop control.

V. Simulation of single cable and multi cable cranes have been provided. Developed control is applicable only for cranes with parallel multi cables.

VI. Suitable reliable open software microprocessor unit Octagon has been selected for the implementation. Tests on real crane were provided.

VII. Design of cable length sensor, implementation of communication with crane driver, etc. were done.

4.2 Semi-Active Truck Suspension [8, 9, 10]

I. Decreasing road-tyre forces can either reduce the road damage or can increase the useful truck payload. The costs of road damage are very high. At the end of the project it was found that the road damage could be decreased by up to 70% or the payload increased by up to 1 ton for 10 tons of original payload.

II. Active and semi-active suspensions, the reasons and cases of their differences have been studied. Different solutions of the conflict between the comfort and road friendliness criteria (Pareto optimization, structural decomposition) have been considered. Preview sensors were investigated.

III. The road tyre forces are proportional to the tyre deformation and can be decreased by decreasing of these deformations. Possible ways are introduction of fictitious damper acting between the tyre and the ground of road similarly as the sky hook (new principle of ground hook), by fictitious cancelling of tyre stiffness (extended ground hook). The conflict with comfort can be solved by the combination with sky hook principle.

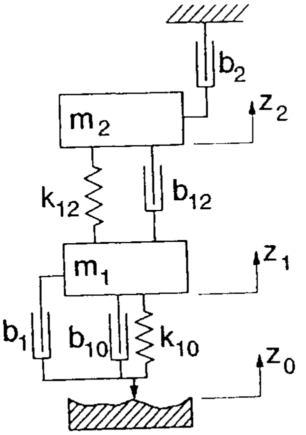


Fig. 1

We are working with fictitious dampers b_1, b_2 and stiffnesses (Fig. 1) the damping forces of which are for real implementation applied on controllable shock absorber F_d (Fig. 2).

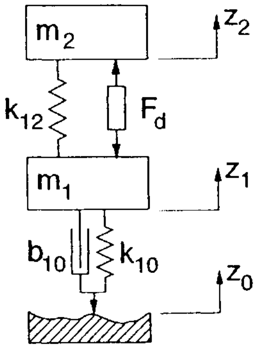


Fig. 2

IV. The actual coefficients of the control laws were set by the parameter optimization which enables to take into consideration the nonlinear truck suspension models. The great attention had to be given to the real properties of controlled shock absorber the dynamics of which influences the possible control achievment significantly.

V. Simulation of passive and semi-active truck suspension, the investigation of model verifications, sensitivities and the influence of real limitations of dynamics of delivered controlled shock absorber were provided.

VI. Suitable reliable open software microprocessor unit, sensor signal filtering, etc. were solved and the prototype was successfully tested.

VII. Design modification for fixing of controllable shock absorbers, reliable wiring of sensor and processors, etc. has to be done.

4.3 Redundant Parallel Robot [11]

I. There is not satisfactory amount of applications of the robots and manipulators in the industry. One of the main reasons for this fact is the low operational speed of robots and manipulators. The speed restriction is based mainly on the maximum reachable acceleration for the whole velocity interval.

II. Parallel robots seem to be one of the promissing ways for the solution of this problem. Parallel robots have several advantageous properties: (A1) All or almost all drives are located on the basic frame which enables to decrease the total mass being moved with the robot and to use drives of any size without limitations due to restriction on the moving mass and stiffness. (A2) Truss construction of robots leads to the higher stiffness than for serial ones.

On the other hand the parallel robots have also several drawbacks: (D1) The robot workspace is generally smaller than for serial robots. (D2) There are generally undesirable singular positions inside the workspace of parallel robots. (D3) The collisions of robot links are generally ofter for parallel robots than for serial ones.

III. The investigation of possibilities of parallel robots has started from planar variants. The initial variant is on the Fig. 3. During the development of kinematic structure of the parallel robot the following requirements have been taken into account: (C1) maximize the workspace, reachable by the maximum angle extent of platform rotation, (C2) preserved good stiffness properties based on the truss construction, (C3) minimize the occurance of singular and nearly singular positions inside the workspace. The initial idea was to increase the robot workspace by usage of redundant DOFs of robot variants, however the workspace increase was small - unsuccessful variants are on the Fig. 4.

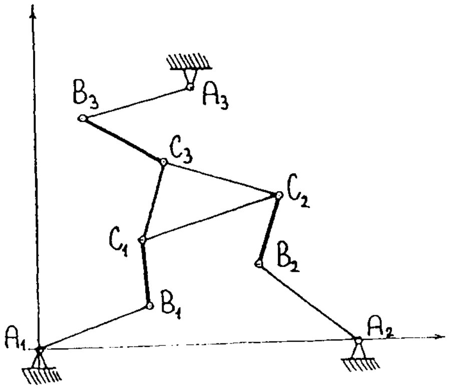


Fig. 3

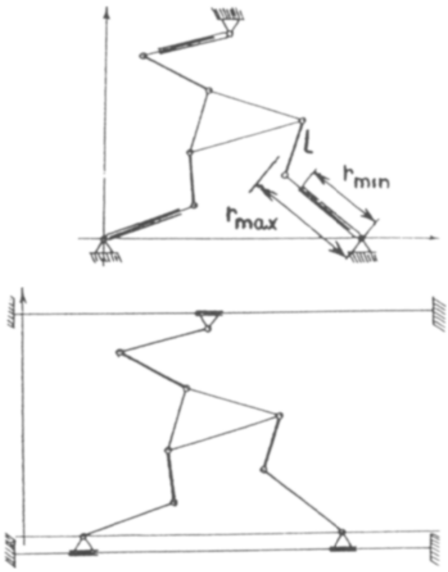
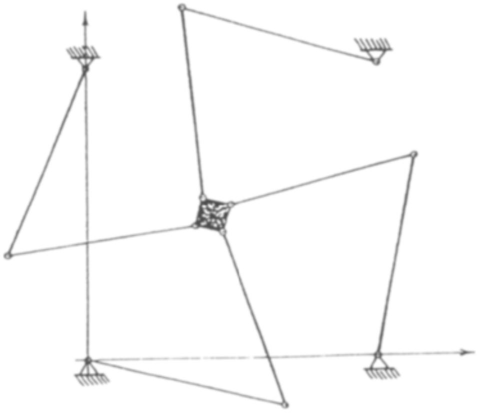


Fig. 4

The deeper and careful analysis has shown that just the suitable change of geometrical parameters can increase the robot workspace and the redundant (redundantly actuated) variant (Fig. 5) is advantageous for removing the singularities and collisions from the robot workspace. This can be interpreted that the search for ideal final solution by conflict resolution was successful and this continued search enabled to overcome the end of design process on the compromises (Fig. 4) before reaching better (ideal) solutions.



IV. Real design of practical robot construction has been proven. However, many new theoretical problems have had to be solved (robot redundancy in drives, robot description by redundant coordinates, optimal control theory for DAE equations etc.)
V. Detailed systematic comparison between suggested redundant parallel robot and traditional singular one has been provided. It has proved that redundant parallel robot has comparable workspace, by one order higher stiffness, four times higher accessible accelerations and two times higher accessible velocities compared to serial robot.
VI. Suitable reliable open software microprocessor unit, driver control design, prototype are being manufactured.
VII. Prototype redesign into the final product is necessary. The derived principles of redundant parallel robots are now being applied for the design of new machining centers.

5. CONCLUSION

The experience from case studies of development process of mechatronical systems and products could be generalized into initial draft of mechatronic design methodology of similar systems (controlled mechanical systems). The summarized key steps of design methodology emphasize the crucial differences to the traditional design steps.

Acknowledgments. The author appreciate the kind support of Czech Grant Agency under the contracts GAČR 101/93/0278 and 101/95/0728 and the kind support of CEC under the contract CIPA-CT94-0130 Copernicus project.

REFERENCES

[1] Valášek, M. (1994): Electrical networks, multibody formalism and mechatronic systems, In: Proceedings of Abstracts, Euromech 320, Multibody systems: Advanced Algorithms and Software Tools, FS ČVUT, Praha 1994, pp. 45

[2] Val ášek, M. (1994): Simulation of mechatronic systems as system with different physical nature. In: Acar, M., Makra, J. and Penney, E. (eds.): Mechatronics: The Basis for New Industrial Development, Computational Mechanics Publications, Southampton 1994, pp. 749-756

[3] Valášek, M. (1994): Inventive engineering as basis of design methodology for mechatronics. In: Acar, M., Makra, J. and Penney, E. (eds.): Mechatronics: The Basis for New Industrial Development, Computational Mechanics Publications, Southampton 1994, pp. 563-566

[4] Valášek, M. (1995): Applicability of Inventive Engineering to Mechatronics, In: Proc. of ICED 95, FS ČVUT Praha, Praha 1995, pp. 1467-1470

[5] Valášek, M., Miláček, S., Dedouch, K., Kozánek, J. and Zolotarev, I. (1996): Position and Velocity Control of Gantry Crane, In: Proc. of Mechatronics 96, Guimaraes, pp. 1-203-1-208

[6] Valášek, M. (1995): Input Shaping Control of Mechatronic Systems, In: Proc. of IFToMM World Congress, Milano, pp.

[7] Valášek, M. et al. (1996): Patent specification PV3991-1996

[8] Novák, M. and Valášek, M. (1996): A New Concept of Semi-Active Control of Truck's Suspension, In: Proc. of AVEC 96, Aachen, pp. 141-151

[9] Valášek, M., Novák, M., Šika, Z. and Vaculín, O. (1996): Extended Ground-Hook -- New Concept of Semi-Active Control of Truck's Suspension, Vehicle System Dynamics, 27(1997), pp. 289-303

[10] Valášek, M., Kortüm, W., Šika, Z., Magdolen, L., Vaculín, O.: Development of Semi-Active Road-Friendly Truck Suspension, IFAC Journal of Control Engineering Practice, 1998 (in print)

[11] Šika, Z., Valášek, M., Miláček, S. and Bastl, P. (1996): Synthesis and Analysis of Planar Redundant Parallel Robot, In: Proceedings of NATO ASI Computational Methods in Mechanisms, Angeles, J. and Zakhariev, E. (eds.), Varna 1997

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

4D Design : adding value to consumer products using mechatronics and multimedia technologies .

Alec Robertson, *MA(RCA) MCSD FRSA*

School of Design & Manufacture, De Montfort University, Leicester. LE1 9BH, United Kingdom.
Tel: +44(0)116 257 7544; Fax: +44(0)116 257 7574; E-mail: alecr@dmu.ac.uk
URL: <http://www.dmu.ac.uk/dept/schools/des-man/4dd/>

Keywords: 4D design; multimedia; mechatronics; product form; value; culture.

Abstract

The common goal of designers is to create an appealing, useful ,useable, and significant product at the right time in the right place for the right customer. In cross-disciplinary design practice, there is at one extreme the ethos of engineering and ergonomic 'functional value', and at the other aesthetics and cultural design with 'emotional value'. The development of microprocessor technology has resulted in the incorporation of computing power into all kinds of 'things'. On the one hand information technology has developed towards a rich multimedia and on the other there has been the integration of control technology through mechatronics design.

This paper discusses some conceptual challenges for adding value to consumer products through the application of mechatronics and multimedia technology in the context of what is becoming known as 4D design, where there is a dynamic cultural component. To illustrate the issue of 4D value the author models different ways for the making of a cup of coffee in relation to the products and people involved.

1. INTRODUCTION

The quality of technical engineering design can increasingly be relied upon. Consumer rights and legal product liability legislation is ensuring the basic characteristics of any product are good - the product will work and do what it is said to do, and it is safe and increasingly environmentally benign. However the 'emotional' value attributed to any product is of crucial importance in the consumer market. Richard Guyatt once Rector of the Royal College of Art, London said " there is such a thing as the ergonomics of the heart'.

Design schools have been instrumental in the growth of professionals, who as well as capable of functional design, are also visually aware about the emotional value of 3D form. Professional design for industry is continually evolving , nevertheless the dominance of the 'visual' has has meant neglect of other form attributes. For example, noise has been tackled but not musical composition. 3D animation as a linear form of narrative entertainment has developed with 'special effects' for blockbuster movies but not the incorporation of kinaesthetics in products - the aesthetics of their movement. Inventions such as the microprocessor have profound implications for design (see Robertson, 1979), and complex dynamic form is one area that now requires attention.

2. 4D DESIGN

The terms *two-dimensional design* and *three-dimensional design* have been used to group certain kinds of designs. They refer to the dimensions of space - the x, y and z axes of geometry. To contrast complex dynamic form with the normally static form of 2D and 3D designs and help cope with the opportunities opening up for dynamic designs, the term *four-dimensional design* is an appropriate addition, where 4D is 'the time dimension' within design .(Robertson, 1992). A definition of 4D Design has been proposed:

> *"the dynamic form resulting from the design of the behaviour of artefacts and people in relations to each and other and their environment." (Alec Robertson, 1995)*

4D design starts with basic animation but it extends beyond it . A football game or play at a theatre or a dance are not 3D animations, and it is this conceptual difference we are dealing with in the context of 4D product design. 3D design draws from knowledge of static art and therefore it follows that since 4D design focuses on dynamic form it incorporates knowledge of the performing arts as well as engineering dynamics.

Real' 4D products have 'dynamic' form in real or actual space . For example this idea is illustrated by the Japanese Tea Ceremony called 'chado',where the cup, table, garments and room are 3D designs and the ceremony is the 4D design. If the teapot on the Japanese tea ceremony 'Chado'' *performed'* dynamically along with the person *participating* in the tea making then it too would be 'real' 4D design. *'Virtual'* 4D products are multimedia designs (see Robertson ,1994) that respond dynamically to their users.

Adding value through 4D design for contemporary consumer products will involve one or both dynamic enabling technologies of mechatronics and multimedia.

Before outlining how 4D design relates to these technologies they are outlined together with a version of CYBERG (Robertson, 1983), a model for assisting product development and evolution.

3. MECHATRONICS AND MULTIMEDIA

3.1 Mechatronics

Mechatronics may be described as an integrative discipline utilising the technologies of Mechanics, Electronics and IT to provide enhanced products, processes and systems. It is a trans-disciplinary approach, based on open communication systems and concurrent practices, to design better engineering products. There have been many attempts to provide tools to aid the mechatronic designer (Yan, T and Sharpe, JE E .1994) and whilst there have been some failures, there is some progress in holistic approaches (Verho, A. and Salminen, V. 1995).

The main area of mechatronics work is in production systems but the miniaturisation of components is leading to more application in consumer products, such as Automatic Teller Machines (ATM) for cash dispensing at banks to video camcorders, and CD players.(Parkin, R.1995) The future has much to offer. The reduction in component sizes is providing new types of actuator and sensors, and with 'smart structures and materials' such as piezoelectric polymers (see Rossi, 1994), is leading to the provision of low cost autonomous robot systems.

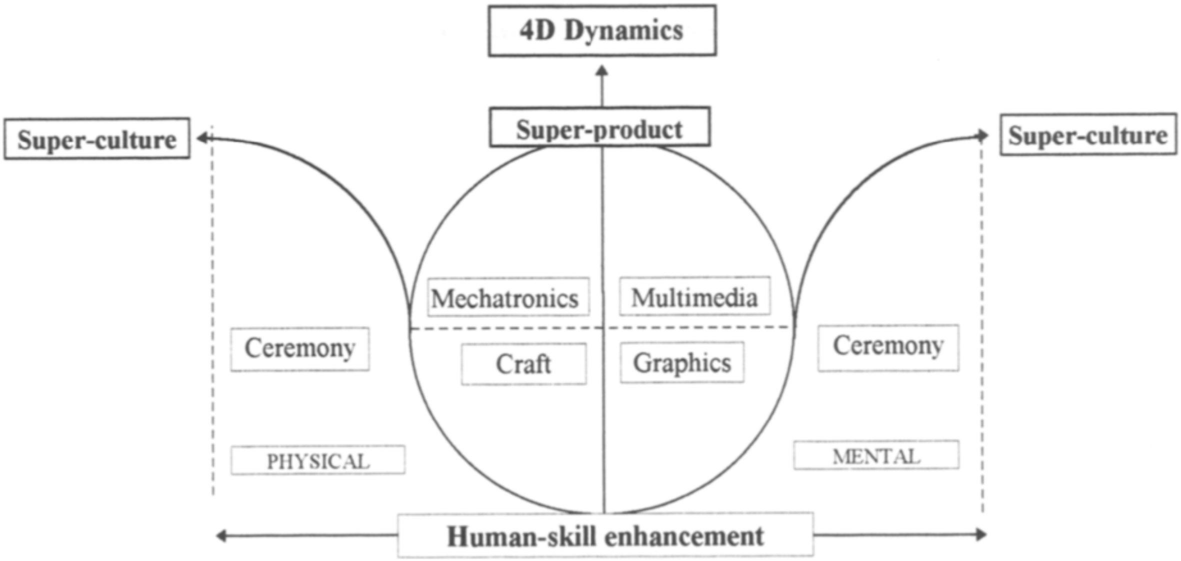
3.2 Multimedia

The integration of media and computer technologies along with the theory of audio- visual information design has led to the growth of multimedia. It enables a more holistic approach to be taken to design of dynamic informational systems .

Application is currently in areas where text, graphics, animation, and sound are useful, such as in teaching and learning programmes eg CD ROM Encyclopaedias, computer shopping information points and the World Wide Web on the internet. Entertainment is another application e.g. computer games and a hybrid term infotainment is used when 'fun' is has useful application. (Sutton, Mark. 1994.)

4. CYBERG

The CYBERG model comprises a framework for creative thought. (See Figure 1). It enables long-term trends in the design and development of products to be taken into account along with the application of new technologies and ideas about the tasks of people and quality of experience people have with products.



.**Figure 1:** CYBERG model for product evolution

CYBERG consists of a form of graph with a horizontal and a vertical axis. The horizontal axis shows levels of human skill-enhancement or replacement in a cultural context, and the vertical axis shows the evolution of 4D form of products. The left-hand side of the graph is concerned with physical characteristics ; the right-hand side is concerned with informational characteristics

Simple products can evolve with ever more sophisticated physical form where more and more physical skill is required by the 'participant' in a human-artefact system. There comes a point when the application of mechatronics technology can be applied to enhance both functional and cultural characteristics.

Likewise simple information designs can evolve with ever more sophistication and there comes a point when information processing power and multimedia technology can be applied to enhance both functional and cultural characteristics.

A human-artefact system may evolve either towards 'super-culture', where there is a high human 'ceremonial' contribution to value, or 'super-product', where there is a high level of autonomy in the artefact.

The crucial question mechatronics engineers to ask is: ' where on the CYB-ERG graph is an established product, and how can it be advanced in its design to add value through 4D design and the introduction of new mechatronics and/or multimedia technologies .

5. 4D PRODUCT DEVELOPMENT

5.1 4D value

Many products we use today seem to have lost 'character' in terms of the distinctive positive features people value. For example the classic SAAB 900 cars are considered by enthusiasts to be more enjoyable to use, with their rich aesthetic of materials, smells and sounds and movements than the new cars. This is clearly not a functional value as they invariable perform less well in terms of speed than newer models. It is not for this paper to discuss why this is but propose how product character can be enhanced through adding value to the 'product experience' by application of mechatronics and multimedia technologies with 4D design.

To begin and show how this can be done an analysis of how the making a drink of coffee or tea can be viewed is decribed with its range of artefacts involved using notions of 3D and 4D design.

It is contended that the evolution of the human-artefact system can take either of the two routes towards - 'super-culture' or 'super-product'. Super-culture involves designing human behaviour through skill and cultural enhancement. Super-product involves diminishing cultural and human skill, but increasing product character and capabilities through intelligent automation with respect for the cultural attributes of a human-artefact system.

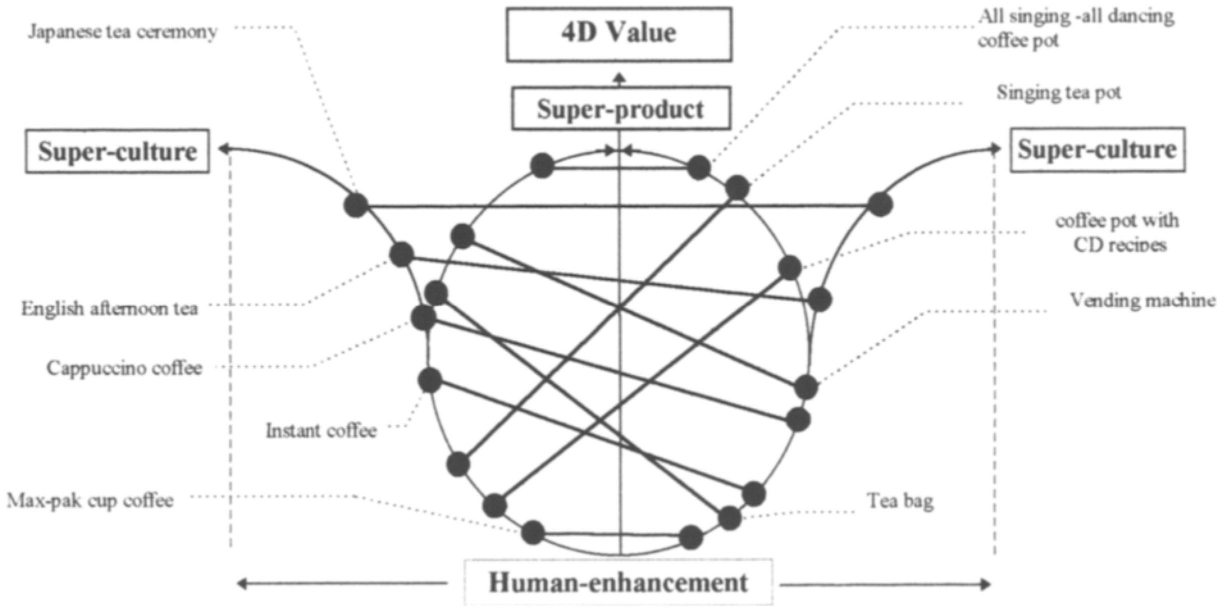


Figure 2: CYBERG 4D

For example, a coffee vending might diminish further in character into where just 'add water', is the only task. Here cold water is poured into a cup with a coffee additive already incorporated, which is heated by a chemical reaction. Alternatively making a cup of coffee could be designed to evolve into something more kinaesthetic, responsive, and dare we say enjoyable. - it would be designed as a cultural 'event'. So how could this be been done using notions 4D design? Figure 2 shows CYBERG-4D and how a variety of coffee designs can be depicted in the 4D context.

5.2 4D designing

The communication of ideas, mood and emotion through action is the bedrock of the performing arts - music, theatre, and dance. This is an untapped source of knowledge for product development and a necessary input for 4D design (Robertson, 1997). The goal of a super-culture would not necessarily require mechatronics and multimedia technologies but the design of new rituals and ceremonies for coffee making. However the goal of creating a super-product would and this is briefly explored below.

A question to ask is 'what characteristics will a 4D product have that a 3D product does not have, and what new knowledge and skills are required?

The answer is found by asking questions with reference to ideas within the performing arts. You could ask for making a cup of coffee, what *'character'* could the products have? For example, would they be friendly? The term user-friendly has been one used by the ergonomist and human-computer interaction specialist. However inspection of 'user-friendly' designs shows they do not really mean 'friendly' but just non-problematic. How would a really user-friendly product *'behave'* ? How would it *'act'*? Could it be given a character to play?

How will the product *'sound'*? Noise is often 'design out' but what dynamic 'audio aesthetics' can be 'designed-in'. The rich sounds of mechanical products have been lost within electronic consumer products so how can designers bring such 'value' back?

How could the product *move*? For example, he technology needed for delivering a paper cup full of coffee from a vending machine is relatively simple, but one can only but notice however that the design of its dynamic form is only functional. Kinaesthetics has not been considered, nor has it had an major influence on the 3D form of the vending machine.

Could the vending machine be made to talk? The art of conversation between people can be a pleasurable experience to be part of. With speech synthesis and voice recognition now on the threshold of practical multimedia application a dialogue with products could be designed and if you are Swedish would you expect the vending machine to speak how you do?

With reference to CYBERG-4D, the Japanese tea ceremony has both a high physical skill and significant cultural meaning, therefore it has a very high 4D value,the ritual English afternoon tea has less super-culture 4D value.

At the other extreme imagine an animatronic all-singing-all- dancing coffee pot using mechatronics and multimedia technologies. This would embody high 4D value but does not require skill enhancement in people. The mechatronics involved would provide imaginative dynamic visual expression in the movement of components.

A tea pot that whistles the tune "Tea for two, and two for tea", when boiling could be viewed as having some 'super-product' 4D value The cappuccino machine requires some skill to use and it has some super-culture 4D value, where the making of your cup in an Italian restaurant is watched with interest. Imagine if a cappuccino machine came with a multimedia CD rom showing different ways to make a coffee in a step by step way. Here the value added would be in the teaching 'performance' of a non-skilled person and therefore will be on the path to super-product.

However a Max-Pac coffee cup which only requires hot water to be added and requires little skill, has a low 4D value of all kinds, but high 3D value due to its static simplicity yet high functional capability. Likewise a contemporary vending machine is technologically quite sophisticated with functional capability but low in 4D value, nevertheless there is an opportunity for advancing vending machine design significantly with mechatronics and 4D design .

Yet the humble simple tea bag which displaces some of the tea making culture could be viewed as diminishing the 'experience' of making a cup of tea and therefore has a low super-culture 4D value, but it transfers value to the products involved with its clever 3D design, which provides a convenient way of measuring tea leaves. It could be that a tea bag design might be developed to encourage cultural dynamics, but this is not for mechatronics engineers.

6. CONCLUSION

The robot is a common theme within films, from Star Wars with R2D2 to the BBC spaceship entertainment series "Red Dwarf" with its 'angry toaster'. The significance of these robots is with the design of their dynamic 'personality'. The functional walking and roving robot has received the attentions of much theoretical engineering research, (see Warwick, 1997), and ergonomics and human computer interface specialists are making an increasing contribution to consumer product design but their input is limited to a scientific and substantially utilitarian approach. There is now a need to explore 4D form as it could be a significant way to add further value for many consumer products incorporating computing power.

It is all too easy to apply a technology such as mechatronics, which functionally 'does' a great deal. However in doing so this may result in unwittingly diminishing the 'super-culture' value'. This may be valid with a corresponding increase in 'super-product' 4D value but otherwise may gain less for the a company's profitablity than could have been the case. The omission of 4D design can explain why some consumer products fail as they do not embody enough consideration of 4D dynamic form.

For example, the experience of shopping is something people enjoy. Much effort is being made to displacing this with on-line multimedia internet sales site, or mechatronically sophisticated vending machines. This should be done with caution, as the 'shopping' experience that people value could be lost.

A strategic approach to design innovation is always required (see Robertson, 1994b), and using notions involved in 4D design could help to reduce risk in product innovation using mechatronics.

Fundamentally, 4D design involves the design of the dynamic behaviour of artefacts and people in relationship to each other and their environment. In other words making the use of a consumer product not only safe and easy but enjoyable too.

Importantly the mix of mechatronics skills with 4D design knowledge and skills in a crsoss-discplinary design team may as well enrich our experience of using consumer products, systems and services in the 21st century.

REFERENCES

Parkin, Rob. (1995) '4D Product Design: A Challenge of Integration'. in 4D Dynamics Conference Proceedings. De Montfort University, Leicester ISBN 1857211306.pp 135 - 148.

Robertson, Alec. (1979) 'Microprocessor Application: Some implications for industrial design'. Unpublished MA thesis. Royal College of Art, London.

Robertson, A., (1983) 'CYB-ERG: A proposed model for assisting innovation in products, and their design'. Creativity and Innovation Network Journal.. Manchester Business School.. Vol. 9, No. 3, July-Sept., 1983

Robertson, Alec. (1992) 'Technolust versus creative design: some implications of intelligent products for design'. Intelligent Consumer Products Symposium. Institute of Electrical Engineers / Chartered Society of Designers. London ,. IEE Digest No. 1992/013.

Robertson, Alec. (1994) '4D Design: The Interaction of Disciplines at a New Design Frontier'. Design Management Journal, Boston USA. Summer 94 pp 26-30.

Robertson, Alec. (1994b) 'Pathfinder products: reducing risk is design innovation'. Proc. 'The International Forum on Design Management Research in Education' 1-3 June 1994. Design Management Institute/ ESCP Senior. Paris, pp 285 -289.

Robertson, Alec (1995), ' 4D Design Futures: Some Concepts and Complexities', in Proc, 4D Dynamics Conference. De Montfort University, Leicester ISBN 1857211306 pp 149 -143 and on CYBERBRIDGE-4D DESIGN WWW site at < http://www.dmu.ac.uk/dept/schools/desman/4dd/>.

Robertson, Alec. (1997) . '4D Product Design: Some conceptual challenges using mechatronics and multimedia technologies'. Proc. PDE 97 National Conference of Product Design Education 7- 8 July. Brunel University.

Rossi, Danilo De. (1994) 'Polymer elecromechanics: mechanical sensing and actuation properties of organic macromolecular systems'. Proc 2nf European Conf. On Smart Structure and Materials. Glasgow, Scotland, pp 19- 26.

Sutton, Mark. (1994) 'To examine the topic of CD-ROM'. Unpublished report, MA Information & Graphic Design Course. School of Design & Manufacture. De Montfort University. Leicester,UK.

Verho, A. and Salminen, V. (1995). Customer Needs, Function Structures and Strategic Development of a Mechatronic Product. Int Conf. on Recent Advances in Mechatronics, ICRAM 95, Aug 14-16 , Istanbul, Turkey.

Warwick, Kevin. (1997)'March of the Machines' Century . London. ISBN 0 7126 7756.

Yan, T and Sharpe, JE E (1994) A Comparative Study of System Simulation Techniques in Mechatronics Proceedings IEE, Control 94, Vol 2 ISBN 0 85296 611 3, pp 1225-1228, 21-24 March, University of Warwick. pp 934-940 .

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

On synergistic aspects in integrated product development of mechatronic systems

F. Kaljas, V. Reedik

Department of Product Development, Institute of Machinery, Tallinn Technical University,
Ehitajate tee 5, EE-0026 Tallinn, Estonia

The aim of the present paper is to contribute to the discussions on mechatronic systems development methodologies. It is pointed out that attaining the maximum synergy of mechanics, electronic control and information technology in mechatronic products is the only real key to enter the highly competitive market environment. Special attention is paid to the use of product development context in the design of mechatronic products. Proposals for the development of the synergistic tools for mechatronic products design are under observation.

1. INTRODUCTION

In 70s the world market of mechatronic products appeared to be full and industry arrived at the truth that in hard competition for world market shares it is possible to survive only by radical cuts in product development time. So different schemes for speeding up products innovation were developed, having parallel, integrated and concurrent patterns of marketing, design, production and financing. Steadily increasing the dynamics and flexibility of mechatronic product innovation, harder price competition, growing concern of customers for product quality, safety and responsibility act as "filters" on the way to the market and have put industry in a very complicated situation. In 90s when the mega-competition age started, the conflict between the growing demands for functionality and quality of mechatronic products at a lower price is the everyday reality in companies and needs continuing renewing of a company's managing structure, work organisation and product development tools to grant the sustainability of the company. Only these companies are successful who are able to road-map technology development ahead up to a decade.

The experience of last decades has shown that product development as a market oriented procedure has no alternative and it is, especially, true for mechatronic products industry where outlined contradictions exhibit themselves even stronger. At the last Engineering Design Conference ICED'97 discussions were held on future mutual relations of engineering design and product development research. It was shown [1] that the design research community is closing to the product development one and there are the following arguments to it as product development technology, compared with engineering design technology, improves design productivity, customer satisfaction, makes lead time shorter and cuts down development costs. So the impact of product development context on mechatronic products design is steadily growing. The power of mechatronics is that the combination of mechanics electronics and software can realise more features than any of these technologies alone. Obviously, it does not happen automatically but only as the result of a systematic and synergistic approach to mechatronic product design.

2. DEFICIENCIES IN SYSTEMS' APPROACH TO MECHATRONIC SYSTEMS DESIGN

The task to design the mechatronic system would be much simpler if the design methodologies of its integrated subsystems were similar, but, unfortunately, it is not so. One of the most comprehensive comparative analysis of mechanical, electronic and software design systems was provided by J. Buur [2]. When he compared such methodological characteristics as functions, concept design, concept realisation, design modelling and methods, they appeared to be quite different. Our repetition of this task ten years later did not reveal signs of coming nearer, it was even vice versa. Until today there exist the following two serious obstacles to integrated design approach:

- abstract functions of systems in mechanical, electronic and software systems are understood and described very differently,
- the means for describing design concepts and conditions of design modelling vary greatly.

Despite a strong impact of mechatronics on intertwining the design approaches the principles on which the initial systems theories are based, are still noticeably different and unevenly developed. One can recognise as the truth that a powerful general evaluation tool to indicate good solutions is still missing. Ultimate decisions here are risky as even the best promising design tool has its own opportunities and shortcomings evaluated from the point of view of functionality, fitness and quality. It is clear that mechatronic system design must be integrated from its early stages.

One of the reasons for deficiencies in mechatronic systems design is probably the division of the mechatronic systems research community into two parts: those who declare that mechatronics is not a separated discipline but a clever mix of a number of different fields of engineering, and those who see mechatronics as an independent discipline with its own outstanding possibilities. But in this mechatronic mix there is obviously "something" more than these technologies offer alone, for example, the realisation of the functions not existing before, increase of flexibility both during design and use (multifunctional ability), compensation of mutual weaknesses (synergy), physical integration, etc. Such a double heart situation may explain the striving of mechatronic research community to concentrate on achievements in real mechatronic systems development rather than general design problems of mechatronic systems. When analysing the last two-three years publications of mechatronics conferences and also library search databases one can notice that also on the basis of the share of publication groups: control and its applications 22%, production automation 19%, sensors and actuators 18%, autonomous guided vehicles 15%, systems modelling and simulation 8%, robotics 7%, artificial vision 5%, mechatronic systems design 4% and mechatronics education 2%.

3. PRODUCT DEVELOPMENT CONTEXT IN MECHATRONIC SYSTEMS DESIGN

Product development philosophy is based on the prerequisite that there is no universally valid method for developing advanced engineering products, but there is a basic generic model of activities, that are necessary for attaining this goal. Integration of these activities with product model development results by dynamic modelling information and communication flows is a real basis of product development. Increasingly important for mechatronic products is market segmentation. "One-size- fits-all" times are definitely over as with technical world complication people have become more different. Mechatronic product development is strongly related with human development towards green design and recycling oriented society. Isolated methods of product design are no longer suitable and integrated approaches are required to involve interdisciplinary and holistic sets of procedures to develop products, considering all phases of the product's life-cycle. Digital product model, product data management system and modular product concept are basic for global marketing through global engineering network (GEN) [6].

All product development tools are equally valuable for mechatronic systems design. Necessary tools can be selected by their functionality, fitness and qualities as some methodologies can be applied throughout the whole design process, some are applicable to individual work steps and phases. There are no obstacles for using QFD (customer voice) for the improvement of the existing products and the same is true about other quality assurance tools – Fault-Tree Analysis, Failure Mode and Effect Analysis, Total Quality Management, etc.. Very suitable at mechatronic systems design are Value Analysis/Value Engineering aiming at maximising the ratio of product function and performance to value of the product. The set of production and operation oriented tools Design for X – production, assembly, quality, reliability, robustness, cost, ergonomics, operation, maintenance, environment, recycling, etc. have been in effective use at mechatronic systems design.

Here we will only dwell upon Design for Reliability, which in mechatronic systems design has a somewhat complicated conception. Against the background of the tremendous progress in information and control technology mechanical parts are likely to be pushed into role of physical "glue". But the mechanical counterpart in mechatronic systems remembers itself through two very important dimensions – in dynamics and temporal behaviour (wear, fatigue). It is necessary to mention that for usual office equipment about 80% of the failures are of mechanical origin and mainly by wear [5]. Despite high-level theoretical approach to reliability problems and strict normative prescriptions for “safety-critical” systems, there are quite a few publications on reliability, safety and responsibility of “non-safety-critical” systems. It is understandable, as in hard competitive environment these data are used only inside the firm. The classical understanding of systems reliability is not very suitable here as besides mechanical, electronic and software failures there are also combined failures on these technologies interfaces (a negative synergy) [5]. To evaluate the optimal level of reliability in product development context it is necessary to determine the responsibility, working speed and price levels needed. It is especially important in the conditions of growing product development frequency, as there is no time to complete the reliability test and the customer is to be involved in follow-up product development.

4. SYNERGISTIC TOOLS FOR MECHATRONICS PRODUCT DEVELOPMENT

Synergy in the integration of mechanics, electronics and software is the only real key for entering the highly competitive market of mechatronic products where the reduction of the weight, energy consumption and life-cycle cost of a new product with higher quality, safety and responsibility at better dynamic performance are decisive success factors. Synergy is here treated as an effect of suitable integration when the whole is more than the sum of its parts. Mechatronic systems are distinguished from classical systems by a greater number of subfunctions, interaction of heterogenuous subsystems and special functional structure. The growing use of the principles of fuzzy control, transputers, neural networks, distributed control cemented by artificial intelligence brings along a new information treatment technology shifting borders in mechatronic philosophy area. They can be configured easily in networking fashion where the execution of the program can be distributed to have the maximum efficiency and speed of performing certain tasks. As a result modules appear which can be assembled in a flexible way to form larger systems and the problem of integrating mechanics, electronics and software crops up both inside the modules and at their integration to bigger subsystems and systems. So an answer is sought to the following question – is the design of mechatronic systems for synergy a miracle or the reality of an endless race with modern technology development.

If you look at the design methodologies on three levels – problem solving in general, the synthesis of technical systems and the total activity of product development, then the first and last are really more universal, not necessarily strongly related to mechanical, electronic and software systems. If you analyse the level of synthesis of technical systems, then you can be convinced that the systems theory and systems thinking are the only possible bases for mechatronic systems design [2,3,6]. This general strategy is based on domains philosophy [7] where the design process consists of the succession – process, function, organ and parts domains with a lot of possible feedback. The function in a mechatronic system is treated as a transformation from input to output forming hierarchical or logical structure. It is clear that the functional structure of a mechatronic system must be described in terms of states and transitions from one state to another. Transformation function is described by the continuous flow of material, energy and information and state functional structures by changes in the mode of functioning. The question how to solve the growing conflict between transformation and state functional structures in the same design models is the most important issue. In other words, we can distinguish logical behaviour of electronics and software and physical behaviour of a mechanical system. For modelling the behaviour the interaction and state diagrams can be used in technology independent manner and the additional domain should be set up between the process and functional domain [8]. In [2] the process and state diagrams are combined also to describe the behaviour of a mechatronic system in realisation-independent manner, indicating under which conditions transitions between transformation processes occur. The following synthesis is based on function/means tree (law of vertical causality) where function is determined as purpose one and a means for realising it include a solution principle (technology allocation). So the next sub-functional level in function/means tree is based on process consideration (horizontal causality). In both last cases the synergy of interaction is not the goal but casual intuitive result.

Impelled by the systems theory approach we have reached close to the understanding that a mechatronic system can be treated as a metasystem, having a higher logical type and viewed from a larger perspective. It is above individual systems, is more than just a collection of individual systems and is a cognitive source of synergy. In [9] a metamodel, applicable to both physical and functional design levels as well as to behavioural topological design descriptions in all energy domains is presented.

In publications an overwhelming trend to modelling and simulation of designed mechatronic systems can be noticed. The mentioned approach is rather analysing-optimising than a synthesis character but at the same time it allows to enfold the dynamic behaviour of mechatronic systems components. To support the basic task of modelling mechatronic products for their synthesis a model description language on three levels was developed [10]. It consists of a dynamic system structure language for topological description on the physical level, a dynamic system language for a block oriented mathematical description on the physical and functional levels and an abstract dynamic system code for machine oriented encoding having metacode character.

On the basis of the analysis of the recent publications it seems that the general theory of mechatronic systems to support the design procedures and principles is so far not available on a descriptive level. But a lot of authors offer prescriptive statements reflecting experience derived from design the process itself.

5. CONCLUSION REMARKS

The truth is that in the modern highly competitive environment only those mechatronic products are successful, in the development of which the synergy of integration of a mechanical system, electronic control and information technology has been achieved. In most cases this success has been more intuition-based than the result of systematic approach. Despite the intensive research of mechatronic systems design, modelling and simulation the power of mechatronic approach does not seem to have been used in full extent so far. It seems that the future benefits from a tremendous progress in information technology will give a new impact on design for synergy, putting the present creative chaos gradually in order. Change in engineering culture, replacing previous competition by new forms of cooperation through global engineering networking may give a good impact on these developments.

REFERENCES

1. T. Tomiyama, Note on Research Direction of Design Studies, Proceedings of the 11th International Conference in Engineering Design ICED'97, Vol.3, Finland, Tampere (1997) 29.
2. J. Buur, A Theoretical Approach to Mechatronic Design, Institute of Engineering Design, Technical University of Denmark, 1990.
3. P. Ringstad, A Comparision of Two Approaches for Functional Decomposition - the Function/Means Tree and the Axiomatic Approach, Proceedings of the 11th International Conference in Engineering Design ICED'97, Vol.2, Finland, Tampere (1997) 57.
4. V. Salminen, E. Buckley, P. Malinen and A. Sauer, Global Engineering Networking – Turning Engineering Knowledge into an Accessible Corporate Asset, Proceedings of the 11th International Conference in Engineering Design ICED'97, Vol.1, Finland, Tampere (1997) 165.
5. T. Tähemaa, On Reliability Criteria of Mechatronic Systems, Finland, Acta Universitatis Ouluensis, Technica, C109 (1997) 205.
6. V. Salminen and A. Verho, Systematic and Innovative Design of a Mechatronic Product, Mechatronics, Vol. 2, 10 (1992) 257.
7. M. M. Andreasen, Syntesemetoder paa system grundlag. Diss. Lunds Tekniska Hogskola, Sweden, 1980 (in Danish).
8. A. Shakeri and R. Braek, Modelling Behaviour of Mechatronic Systems, Proceedings of the 11th International Conference in Engineering Design ICED'97, Vol.3, Finland, Tampere (1997) 741.
9. H. Mann and H. Van Brussel, Metamodel and Design Methodologies for Mechatronics,. Proceedings of Tampere International Conference on Machine Automation – Mechatronics Spells Profitability, Finland, Tampere (1994) 671.
10. J. Richert and M. Hahn, DSS-DSL-DSC – the Three Levels of a Model Description Language for Mechatronic Systems, Proceedings of Tampere International Conference on Machine Automation – Mechatronics Spells Profitability, Finland, Tampere (1994) 777.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Design, fabrication, and testing of miniature polymer gripper systems

R. J. Chang[a], H. S. Wang[b], and Y. L. Wang[b]

Department of Mechanical Engineering, National Cheng Kung University,
Taiwan, Tainan 701, Republic of China

[a] Professor
[b] Graduate student

A prototype miniature gripper system with mechanism made of polyurethane (PU) is designed, fabricated, and tested. The gripper system including gripper mechanism, actuator subsystem, and sensor subsystem is designed. Gripper mechanism is fabricated and loading-deflection behavior is measured. The mechatronic gripper system is finally integrated and tested. The gripper dynamics is identified and the system specifications are listed.

1. INTRODUCTION

Miniature grippers that are capable of handling small objects have many important applications in information industries [1]. Considerable attention has been given to the design and fabrication of miniature gripper systems for micromanipulation during fabrication process, assembling operation, and maintenance works. The stringent technological requirements of small size, clean operation, fast response, and high accuracy are cost-ineffective in conventional design of gripper systems. Recently, the researches on microelectromechanical system (MEMS) and compliant machine have stimulated the interests on compliant microgripper systems to achieve the technological requirements [2,3].

Compliant gripper systems, by following the general classification of compliant machine, can be classified into two categories [3]. One is a system designed and fabricated through the assembly of compliant mechanical elements. Another is designed and fabricated as a one-piece no-assembly system with output motion generated through compliant operation. With one piece no-assembly gripper design, many critical specifications such as linear displacement, micro accuracy, zero backlash, no friction, and no wear can be achieved through exploiting compliant design in micro operation. In addition, one-piece gripper design is easily miniaturized and manufactured from employing extrusive or moldable material.

Design and fabrication of miniature gripper systems involving sensor, mechanism, actuator, and controller for critical operations is a challenge work for mechatronic engineer [4]. Since the first silicon microgripper was developed and proposed by Kim etc. in 1990 [5], several prototype miniature grippers have been designed and fabricated. Microgrippers, in general, can be classified into two categories. One is a gripper itself being around micro size and fabricated by MEMS technology [6]. Another is a gripper for handling micro objects and fabricated by precision machining technology [7]. The manufacturing technologies for mechanical subsystem with silicon or metal material to achieve micro size or operation are rather expensive. In considering the cost effectiveness and high manufacturibility, polymer material is selected for the fabrication of present miniature gripper mechanism.

In this paper, a prototype miniature gripper system with mechanism made of polyurethane (PU) is designed, fabricated, and tested. System design including conceptual design and subsystem design will be first described. Then, the paper will be devoted to the characterization of the mechanical properties of PU and the fabrication process for the implementation of the miniature gripper previously designed. The next part of the paper concerns experimental test of the gripper system. Experimental setup employed for this is described and measurement results are presented and discussed. Finally, the present research on the prototype miniature gripper system will be concluded and further development will be recommended.

2. SYSTEM DESIGN

2.1. Concept design

The present design is to follow Suh's design axioms [8] on mechanical, electrical, and control subsystems for realizing a mechatronic gripper system. The mechanical subsystem including end-effector, transmission, and suspension is designed and fabricated as a one-piece mechanism from employing extrusive or moldable polymer material. The operation of gripper is actuated by electromagnetive force through the deflection of compliant gripper mechanism. The electrical subsystem including signal detection, impedance matching, and power amplification is implemented by employing bridge circuit, isolator, and power transistor. The controller including compensation, feedforward, and feedback functions for the control signal of contact detection, contact friction learning, and grip holding will be implemented by employing C language and running on a personal computer.

2.2. Subsystem design

2.2.1. Gripper mechanism

Compliant microgrippers are designed to provide required tip motions and gripping force transmission through their deformation under the action of actuating forces. For such mechanisms employed by exploiting compliant deformation in micro operation, proper topology, optimal shape, and accurate size are necessary design consideration to provide micro, accurate, and effective output performance. As with type synthesis in conventional gripper design, topological synthesis is the first phase undertaken in the synthesis of compliant grippers.

For the first phase of designing a lumped compliant gripper, a proper topological structure is selected from the existing types of conventional gripper mechanism. Then, the selected gripper mechanism is transformed into a PLM (pseudo-linkages model) with equivalent bending springs [9]. By utilizing an atlas of sixty-four gripper topologies listed by Belfiore and Pennestr [10], a six links and six joints design is selected under the consideration of symmetric gripper arms, simple elastic and rotational joint, and linear drive. In order to obtain a single degree-of-freedom operation of the selected mechanism, the two gripper arms are assumed to be driven synchronously by a sliding mechanism. The second design phase is to synthesize an optimal shape and size for the configuration of a compliant gripper. By employing the design constraint of planar area of 10mm by 10mm for the gripper mechanism, an optimal PLM configuration is obtained under the design objective of maximum linear displacement gain and minimum rotational gain of tip motion. The detail geometry and size of the compliant microgripper is finally determined under the consideration of manufacturing technologies, stiffness of compliant joint, and availability of polymer material. The final layout of the present design of polymer gripper mechanism is shown in Figure 1.

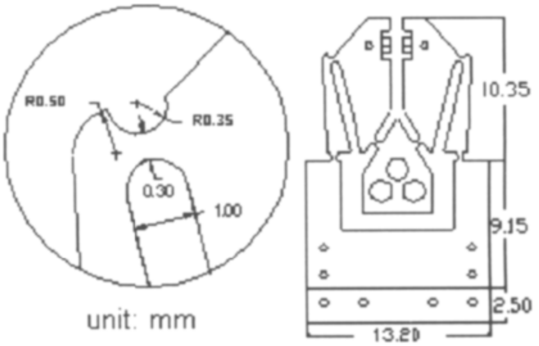


Figure 1. Layout of gripper mechanism

2.2.2. Actuator and drive [11]

Actuation subsystem is designed to provide efficient and effective operation on driving gripper mechanism. Electromagnetic actuation including electromagnetic actuator and electrical drive usually can be designed either by driving a coil or a magnet. With the requirement of low mass on the moving platform of a gripper, a dc moving-coil linear actuator is selected. The driving force of actuator is proportional to both the flux density by magnet across a gap in which the driven coil is positioned and the strength of the magnetic field produced by the coil.

By analyzing the flux density across the gap, it is realized that the maximum flux density can be obtained for a magnetic material having the highest energy product in operation and the minimum gap volume under size constraint. The moving coil actuator is with drawback of heat generation when electrical power is supplied. Minimization of heat generation by proper design of coil winding and power drive is required.

2.2.3. Sensor and signal processor

Measurement subsystem including sensor and signal processor is designed to provide sufficient and reliable information of gripper and environment status for the control of gripper operation. The operation of mechatronic compliant gripper system can be controlled through measuring the signal of position, force, stress, strain, current, voltage, magnetic flux intensity, and magnetic field. In considering the phase of prototype development, an opto-electronic position sensor is selected for on-line motion measurement and control of gripper tips.

3. GRIPPER FABRICATION AND MEASUREMENT

3.1. Measurement system

In order to obtain physical data of loading and deformation of planar motion, a measurement and testing system was installed and calibrated. The measurement and testing system consists of three subsystems: loading platform, image system, and data processor with display. The loading platform is composed of optical table, x-y stage, and loading direction regulator. The image system includes SONY CCD XC-75 (Image Frame Grabber), monitor, and EA4 0.1 160 lens. The lens in the image system is with 34mm in focal length and 51mm in object distance. The image resolution is 5μm for each pixel with 640x480 pixels for a frame. In the data processor, a micro computer is used to convert image data to real displacement.

3.2. Fabrication process

The prototype gripper mechanism is fabricated under the consideration of gripper design, polymer material availability, and manufacturing ability. With the gripper mechanism designed previously and the polymer material which is available from local industry, the miniature gripper with thickness of 1mm is fabricated by using PU material with hand tools. The hand tools including a set of files, cutting knife, drills, and hammer are employed in fabrication process. The fabrication sequence from layout design to final product is shown in Figure 2.

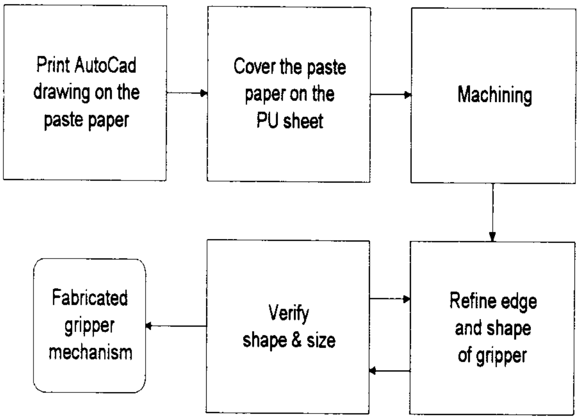

Print AutoCad drawing on the paste paper
Cover the paste paper on the PU sheet
Machining
Refine edge and shape of gripper
Verify shape & size
Fabricated gripper mechanism

Figure 2. Fabrication sequence of gripper mechanism

3.3. Loading and relaxation test

The material properties of Young's modulus (E) and Poisson's ratio (υ) of PU are estimated from experimental data under loading-deformation test. The specimen with overall length of 10 mm has wedge-shaped ends with a narrow central test section. Two tiny reflective indentations with 1 mm apart for deformation measurement are impressed in the specimen surface. By performing loading test to obtain the deformation data, the material properties with $E=1.3085\times10^{7}$ Pa, $\upsilon=0.4669$ can be estimated from linear elasticity theory.

The loading and relaxation behavior of the gripper are measured by experimental test. Measurement procedure for various physical behavior of the miniature gripper is undertaken. In this experiment, a loading weight is connected to the gripper with a wire of 0.2mm in diameter. The deformation of gripper tip is measured by the CCD camera of measurement system.

The creep deformation and relaxation of gripper mechanism will be first tested. The response of gripper tip by 0.095gw load can level off in almost thirty seconds. The tip deformation will not return to initial unloading position immediately until wait for thirty seconds later. Next, cyclic loading and unloading tests are proceeded. For each cyclic test, a stepwise load of 0.095 gw is added or released every thirty seconds after tip deflection is almost fully developed. The tip deformation to plunger displacement and input load applied are measured, respectively, to give Figure 3 and Figure 4.

Experimental results from Figure 3 reveal that a linear displacement gain of 2.6611 mm/mm can be obtained. From Figure 4, it is seen that the first two loops are always inconsistent with the other loops and should be ignored in modeling the loading-displacement relation. The high hysteresis band gives the advantage of high passive damping in gripping operation. By taking the average of three loops of loading and unloading curves and employing least-square data processing, a mean force to displacement relation can be estimated. The estimated relation is with 64.03 mm/N gain and 0.04 mm bias. The bias which is not consistent with linear elasticity theory needs to be compensated in later model development.

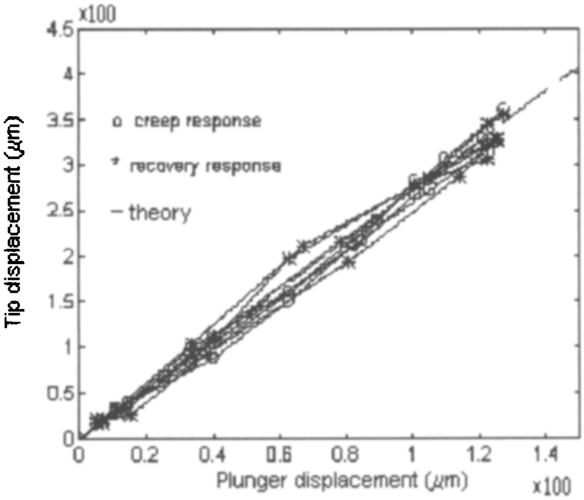


Figure 3. Plunger displacement to tip displacement

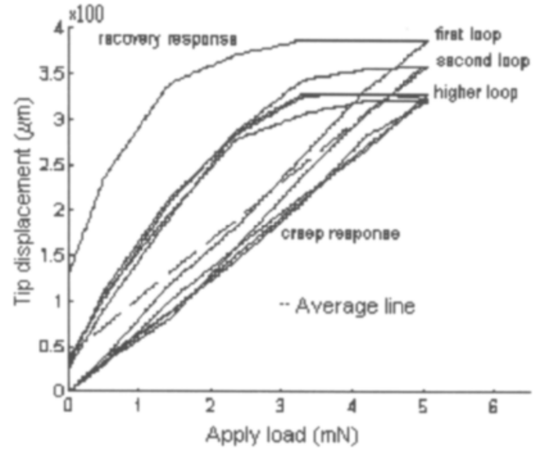


Figure 4. Applied load to tip displacement

4. SYSTEM TESTING AND SPECIFICATIONS

4.1. Experimental setup

The experimental setup is shown schemetically in Figure 5. The fabricated miniature gripper system can be mounted on a testing platform with step driving motor for horizontal and vertical motion driven testing. The overall system is controlled by utilizing 286 PC.

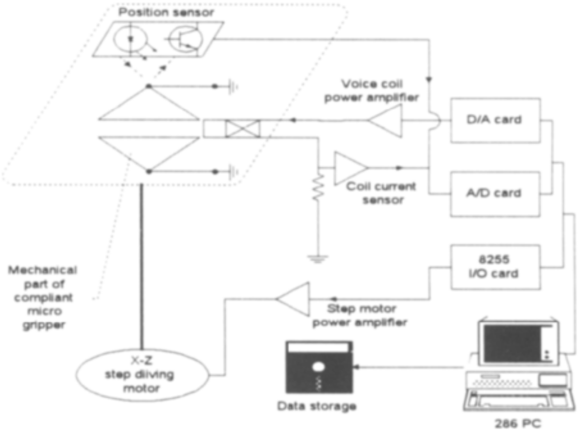


Figure 5. Experimental setup for system testing

4.2. System integration

The polymer gripper system is fabricated by the assembly of mechanical, electromechanical, and electrical subsystems. The gripper mechanism and coil winding previously designed and fabricated are first assembled by simple hand tools with adhesive glue. Then, the permanent magnets are assembled on mechanical frame by using holding plates and screws. Finally, an opto-electronic position sensor, KT3092, which is avilable from industrial products, is assembled. The fabricated prototype miniature gripper system is shown in Figure 6.

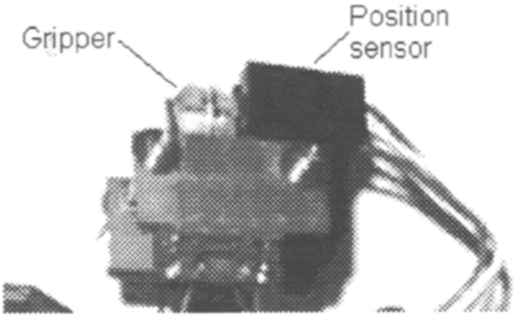


Figure 6. Photograph of polymer gripper system

4.3. Mechatronic system test

The micro mechatronic gripper system is tested and measured to obtain characteristic curves of sensing, actuation, and gripping. The gripper tips are tested by applying step input voltage and the resultant deflection is measured from the opto-electronic position sensor. The step response of gripper system is shown in Figure 7. The Bode plot estimated from impulse response is shown in Figure 8. From Figure 7, it is observed that settling time of 50ms can be obtained. Figure 8 reveals that the resonant peak of gripper system is located around 700 rad/sec (110 Hz). The mechatronic gripper system is identified to give transfer function from input voltage (V) to output tip displacement (m) as:

$$\frac{X(s)}{V(s)}=\frac{65755}{s^3+408s^2+596181s+208492144} \quad (1)$$

The validity of the transfer function is supported by step testing and Bode plot as shown in Figure 7 and Figure 8, respectively. The general system specifications are listed in Table 1.

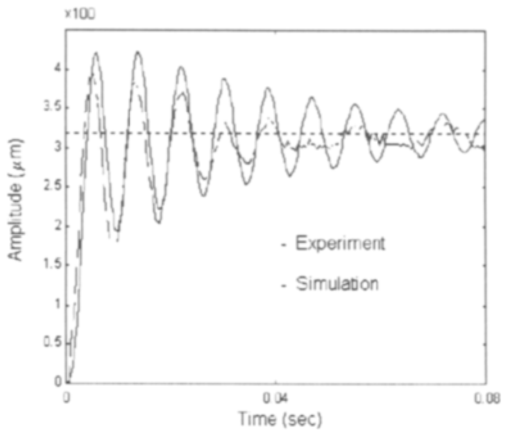


Figure 7. Step response of gripper system

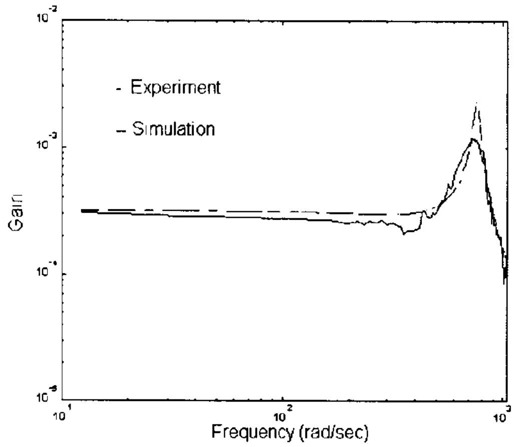


Figure 8. Bode plot of gripper system

5. CONCLUSION

One-piece polymer gripper mechanism is easily miniaturized and manufactured to achieve high performance with low cost expense. The manufacturing accuracy is realizable with the gripper size around $1\,cm^3$. The operational accuracy of mechanism is around 5 μm. Voice coil actuator is sufficient to provide actuating force of bilateral direction with the gripping range up to 2 *mm*. In the gripping operation, a normal load with 0.5gw can be held under open-loop control. The mechatronic gripper system is with 50ms settling time. Implementation of neural-networks friction learning and closed-loop control will be further developed.

Table 1

Specifications of micro compliant gripper system

size	11X10.35X1 mm^3 (gripper mechanism only) 13.2X22X14 mm^3 (whole gripper)
weight	0.274 gw (gripper mechanism only)
material	PU
actuator	voice coil 2 layer 54 turns 5X4X2 mm^3
sensors	1.opto-electronic position sensor 2.voice coil current sensor
displacement gain	2.6611 mm/mm (tip/plunger)
force gain	0.3758 N/N (tip/plunger)
equivalent compliant constant	64.03 mm/N(tip/load)
rating gripping force	0.558 gw
peak gripping force	0.866 gw
max. gripping size	2 mm
settling time	50 ms
resonant frequency	110 Hz

ACKNOWLEDGMENTS

The authors want to thank the NSC of Taiwan for the support under contract No. NSC87-2212-E-006-022.

REFERENCES

1. I. Fujimasa, Micromachines, A New Era in Mechanical Engineering, Oxford University Press, New York, 1996.
2. S. Kota, G. K. Ananthasuresh, S. B. Crary, and K. D. Wise, J. Mech. Design, Trans. ASME, 116 (1994) 1081.
3. A. Midha, T. W. Norton, and L. L. Howell, J. Mech. Design, Trans. ASME, 116 (1994) 270.
4. D. A. Bradley, D. Dawson, N. C. Burd, and A. J. Loader, Mechatronics, Electronics in Productions and Processes, Chapman & Hall, 1991.
5. Chang-Jin Kim, A. P. Pisano, R. S. Muller, and M. G. Lim, ASME, DSC, 19 (1990) 99.
6. G. K. Ananthasuresh, S. Kota, and N. Kikuchi, ASME, DSC-2, 55 (1994) 677.
7. L. Sun, J. Chen, and P. Sum, 6th Int. Symp. Micro Mach. Human Sci., (1995) 199.
8. N. P. Suh, The Principles of Design, Oxford University Press, New York, 1990.
9. R. J. Chang and Y. L. Wang, 14th Nat. Conf. Mech. Eng., CSME (1997) 23.
10. N. P. Belfiore and E. Pennestri, Mech. Mach. Theory 32 (1997) 811.
11. A. P. Dorey and J. H. Moore (ed.), Advances in Actuators, Institute of Physics Publishing, London, 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Design Methodology of Application Specific Integrated Circuits for Mechatronic Systems[1]

F.-M. Renner[a], K.-J. Hoffmann[b], R. Markert[b] and M. Glesner[a]

[a]Institute of Microelectronic Systems, Darmstadt University of Technology,
Karlstr. 15, 64283 Darmstadt, Germany.

[b]Institute of Mechanics, Division of Dynamics, Darmstadt University of Technology,
Hochschulstr. 1, 64289 Darmstadt, Germany.

In this paper we describe the implementation of a controller for mechatronic applications. First, a brief overview of the mechatronic application, an integrated magnetic bearing system, is given. The design process for a software and hardware prototype implementation is described and details of the individual modules within the design process are given. The controller of the magnetic bearing is implemented using a graphical high-level approach. From this implementation a C/VHDL description is automatically generated and transferred either as a software prototype to a DSP system or as a hardware prototype to a FPGA based system located at the test rig.

1. INTRODUCTION

The traditional designflow of mechatronic systems starts from the specification of the end product. Based on a first concept, the system is divided into a mechanical and an information processing part. After defining the interfaces for the communication between these system components, the development for the mechanical and electronical part is carried out independently. At the end of the development cycle the integration takes place and the system can be validated against requirements. The main disadvantage of this approach is that the test for validation is done at the end of the development process. Errors made in earlier design stages force redesign cycles which are expensive in terms of Time-to-Market and Design-to-Cost.

In the proposed design methodology, starting from system requirements and information processing algorithms, dataflow graphs are captured in a high-level block-diagram based, visual environment. After verifying the algorithm the designer has the choice to rapid-prototype the system using common-off-the-shelf (COTS) components, to emulate the system based on field programmable components, like FPGAs or to develop an application specific integrated circuit (ASIC).

The physical realization of the controller can be done in two ways (see Figure 4). In a first step after the validation of the magnetic bearing system, a software prototype is generated in a matter of seconds and transported to the target hardware located at the test rig. This fast generation of software prototypes enables the designer to modify the algorithm used. Furthermore different control algorithms can be tested under realistic conditions in order to select the optimal algorithm for the magnetic bearing system, [1], [2]. Besides the C code generation, VHDL code can be generated from the initial specification as well. This hardware description language enables the designer to map the functionality of the controller on a FPGA that can be reconfigured if necessary. Finally an ASIC can be fabricated and put on a printed circuit board within the power amplifier.

2. THE INTEGRATED MAGNETIC BEARING SYSTEM

The design methodology described in this paper is applied to realize an integrated active magnetic bearing system. Conventional magnetic bearings

---

[1] This project is funded within the special research program 241 (IMES) by the German Research Foundation (DFG).

consist of a complex assemblage of individual components of widely differing functionality, each of them performs its own unique task. To simplify the active magnetic bearing system it is essential that the individual components are integrated both functionally and structurally, as far as possible. The result is a self-contained device with a considerably improved capability operating as a "black box" which is externally controllable. Furthermore, the cost of manufacturing and operation will be lower than they are at the present. This will be achieved for example by including nonlinear modeling, sensorless determination of the rotor position, design of adaptive control schemes, and the implementation of the software in an application specific integrated circuit (ASIC).

Here we will consider the compensation of the nonlinearities of active magnetic bearing systems by the controller software and the integrated circuit design necessary for it. Most components of magnetic bearing systems are non-linear, thus making the entire system non-linear, [3].

The non-linearities, usually neglected for computational simplicity, become strongly evident for large currents, high frequencies, large magnetic forces and small air gaps and restrict the working range of magnetic bearings. So, for example, the softening force-current characteristic of the electromagnets and the limitation of the coil currents by the power amplifier yield resonance curves bending to the left. This can result in undesired jumps of the rotor-amplitude during run-up and run-down accompanied by transient vibrations.

The essential non-linearity of active magnetic bearings results from the force-current-displacement-characteristic of the electromagnets. Usually high pre-magnetization currents $i_V = i_{max}/2$ are used to reduce the effects of this non-linearity on the dynamic behavior of the rotor.

Yet, only a rough linearization can be achieved by this method. As an alternative the system can be linearized by software using a two-component controller (see Figure 2).

The controller consists of a conventional part and a post-connected compensator that *inverts* the magnetic force characteristic with regard to the current. The measured force characteristic (Figure 1) should be approximated by a mathematical formula which can be inverted.

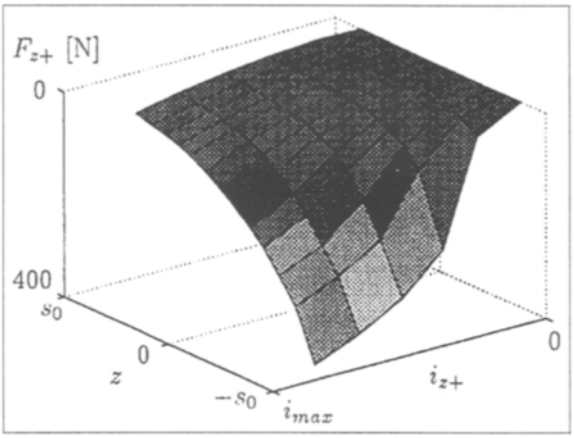


Figure 1: Measured force characteristics

Their inversion is taken as compensator function and post-connected to the conventional controller part. The compensator determines a corrected control variable $u_{Kz+} = f_K(u_{z+}, z)$ depending on the previous control variable u_{z+} and the rotor position z so that the system as a whole is linear ($F_{z+} = k_{Fu}u_{z+}$). The force-displacement factor k_{Fu} can be chosen arbitrarily and for certain f_K the magnetic force even becomes independent of the rotor deflection z for constant currents. The inverse characteristic as well as the functionality of the compensator is stored on a micro computer to that the superordinate controller can access on-line.

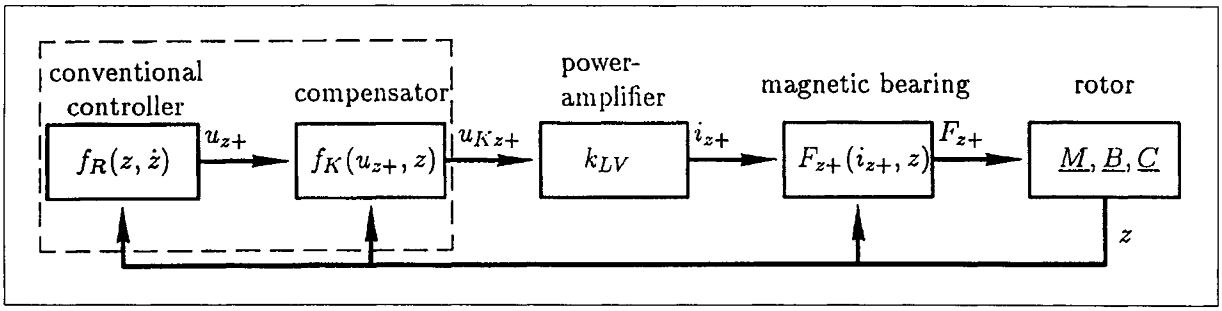


Figure 2: Control loop of an active magnetic bearing with compensator

The linearization by software was tested in simulations as well as on a rotor test rig, [4]. With activated compensator the non-linear characteristic is compensated and the rotor system behaves linear.
In simulations even the extreme case $i_V=0$ was possible. But in practice the total reduction of the premagnetization currents to zero does not work because the current's slope is limited by the power amplifiers. The exact description of the process and the characteristic of the magnetic force is the crucial point for the success of this functional integration.

3. HIGH LEVEL SPECIFICATION

The information processing part of an integrated magnetic bearing system consists of different representations. On one side there is a control algorithm part, which is a dataflow oriented representation of the control algorithm, on the other side there is a controlflow oriented part, realizing for example the communication with the outside world and the configuration of the system after reset.
Figure 3 shows the general designflow considered in our work. Starting from system requirements and information processing algorithms, dataflow graphs for the information processing part and finite state machine models for the control part are captured in high-level visual environments.

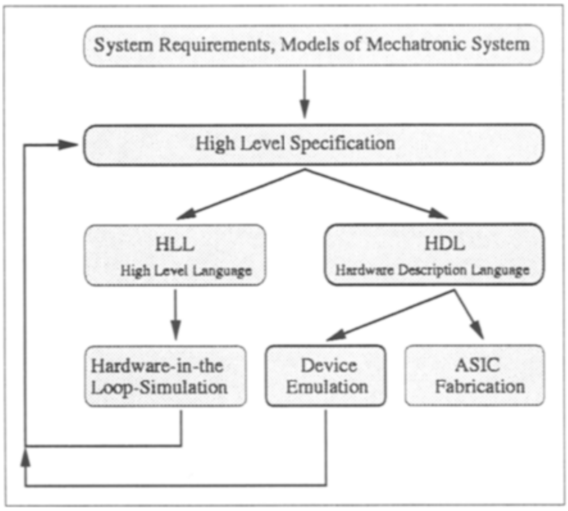

System Requirements, Models of Mechatronic System
High Level Specification
HLL
High Level Language
HDL
Hardware Description Language
Hardware-in-the Loop-Simulation
Device Emulation
ASIC Fabrication

Figure 3: Designflow for mechatronic systems

Starting from the high-level specification of the algorithm for information processing, our goal is to turn around such designs for mechatronic systems in a few weeks, focusing on common-off-the-shelf (COTS) components, [5], [6].
The wide range of different algorithms to be considered for magnetic bearings makes it impossible for the designer to know in advance the characteristics of the system to be developed in detail. However, for these information processing algorithms, synchronous dataflow has been proved to be a natural representation, [7].
One of its strengths is the exposing of parallelism by expressing only the actual data dependencies existing in the algorithm. Dataflow is a well-known programming model in which a program is represented as a directed graph, where the vertices represent computation and edges represent FIFO (first-in-first-out) queues that direct data values from the output of one computation to the input of the following computation. Data precedence between computations are thus represented by edges. Actors consume data (or tokens) from their inputs, perform computation on them and produce certain number of tokens on their outputs, [8].
The dataflow models implemented within the magnetic bearing application is the SDF - Synchronous Data Flow - model proposed by Lee and Messerschmidt, [9]. The advantage of dataflow graphs is its formalism for block-diagram based visual programming, which is an intuitive specification mechanism for digital signal processing. The functional building block executes a specific algorithm but the eventual implementation depends upon the particular design steps the designer wishes to follow.
For the modeling of the behavior of the controlflow oriented part of the magnetic bearing controller an extended finite-state-machine (FSM) description was created based on Statecharts, [10]. Statecharts are an extension of the traditional FSM description including three additional features: hierarchy, concurrency and communication. This design method was used before to model a CAN-Bus Controller, [11].
Using this level of abstraction lets designers (computer engineers, electrical engineers as well as mechanic engineers) complete the tasks at a very early design stage as efficiently as possible.

4. RAPID SYSTEM SPECIFICATION FLOW

Our approach (see Figure 4) is based on the commercial tools Signal Processing Worksystem (SPW) from Cadence Design Systems and ExpressV-HDL or Statemate from i-Logix.

Starting from system requirements and information processing algorithms, dataflow and controlflow graphs are captured in a high-level block-diagram based visual environment. After verifying the algorithm the designer has the choice

1. to rapid-prototype the system as a software prototype using COTS components, like DSP microprocessors, AD/DA converters, etc., or
2. to emulate the system as a hardware prototype based on FPGA components, or
3. to develop an Application Specific Integrated Circuit (ASIC).

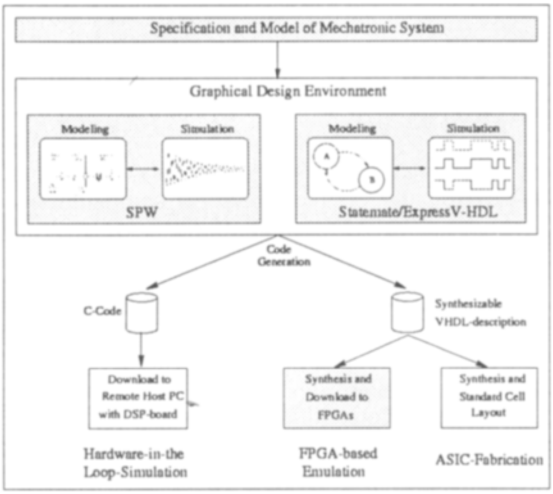

Specification and Model of Mechatronic System
Graphical Design Environment
Modeling
Simulation
SPW
Modeling
Simulation
Statemate/ExpressV-HDL
Code Generation
C-Code
Synthesizable VHDL-description
Download to Remote Host PC with DSP-board
Synthesis and Download to FPGAs
Synthesis and Standard Cell Layout
Hardware-in-the Loop-Simulation
FPGA-based Emulation
ASIC-Fabrication

Figure 4: Our detailed designflow

One of the key aspects in our approach is that the designer is able to start step 2 and 3 from the same representation of the graphs instead of having to re-specify and reevaluate the system specification into a lower level hardware description language like VHDL. The dataflow of the algorithm is captured in a graphical editor using functional blocks provided by SPW as well as custom coded blocks. The functionality of the algorithm can be evaluated by simulating the system within SPW. After successful simulation, a C description of the whole design can be automatically generated from SPW targeting a DSP processor. This software prototype can be used for validating the systems functionality. With successful validation of the rapid software prototype, the next step is to generate automatically a hardware prototype. First, a floating point to fixed point conversion has to be performed. After simulation of the fixed point system the VHDL code of the dataflow part can be generated automatically. However, to be able to put this system on a FPGA a supervising controller has to be implemented as well. The controlflow oriented part is modeled with ExpressV-HDL from i-Logix as a hierarchical finite state machine (FSM) using the key advantages of ExpressV-HDL: hierarchy and concurrency. The controller has to initialize the registers upon reset, sets default parameters for the dataflow oriented algorithm and monitors the interface to an external PC if a change of parameters occurs. The functionality of the controller can be graphically simulated within this tool in order to verify it's correct behavior. After successful simulation the VHDL Code of the controlflow oriented part can be automatically generated as well. In this way the complete information processing part for the magnetic bearing system can be implemented using high-level graphical environments which can be used by electrical and computer engineers as well as mechanic engineers without having to consider the actual hardware structure of the final system at this early design stage.

5. SOFTWARE PROTOTYPING

The automatically generated C code is used to create a software prototype. The C code is transferred to a remote PC including the prototyping board located at the test rig.

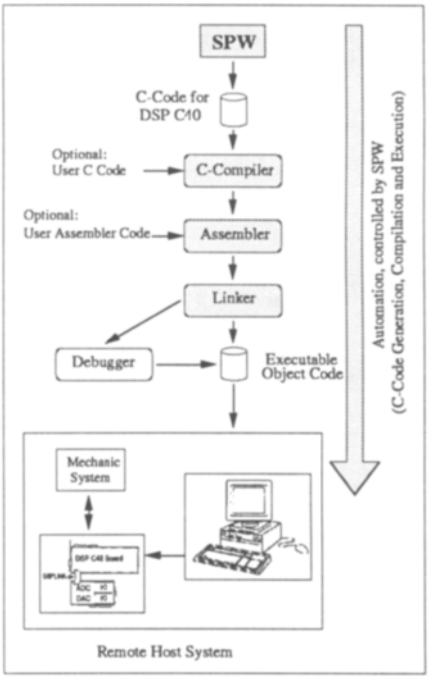

SPW
C-Code for DSP C40
Optional: User C Code
C-Compiler
Optional: User Assembler Code
Assembler
Linker
Debugger
Executable Object Code
Automation, controlled by SPW (C-Code Generation, Compilation and Execution)
Mechanic System
DSP C40 board
ADC
DAC
Remote Host System

Figure 5: Software prototyping designflow

Our DSP C40 prototyping board can support up to 4 modules containing TMS 320C40 processors as well as I/O peripherals. Operating at 50 MHz, one TMS 320C40 achieves a performance of 25 million instructions per second (MIPS) and is thus well suited for mechatronic applications. The analog interfaces to the mechanic system consist of two I/O modules, each of them is a 16 bit dual channel I/O solution for sample rates up to 200 MHz. Figure 5 shows the designflow for a software prototype generation. The C code generated by SPW is transferred via ftp to the remote PC located at the test rig. The C code is compiled, downloaded to the DSP board and initialized. The software prototype can now be validated in a hardware-in-the-loop simulation. Changes of the architecture can be made in SPW in a redesign cycle whereas changing of parameters can be done online.

6. HARDWARE PROTOTYPING

After generation of the VHDL code for the controlflow and dataflow oriented parts of the system the functionality is synthesized on Field Programmable Devices, like FPGAs in order to evaluate the functionality of the information processing part within the dedicated hardware environment.

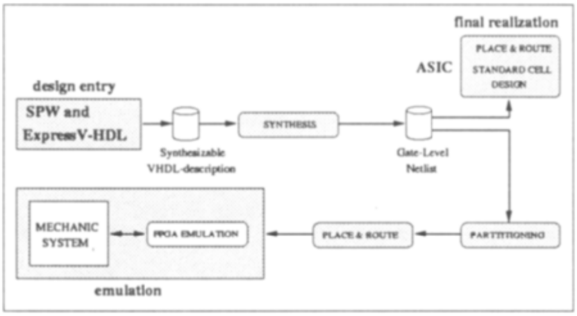

final realization
ASIC
PLACE & ROUTE
STANDARD CELL DESIGN
design entry
SPW and ExpressV-HDL
Synthesizable VHDL-description
SYNTHESIS
Gate-Level Netlist
MECHANIC SYSTEM
FPGA EMULATION
PLACE & ROUTE
PARTITIONING
emulation

Figure 6: Hardware prototyping designflow

According to Figure 6 the hardware prototype of the information processing part for integrated magnetic bearings is generated as follows:

The controller is completely specified using the graphical design environments SPW and ExpressV-HDL (see Section 3). From this specification VHDL code is generated which is fully synthesizable. A standard synthesis tool (Synopsys Design Compiler) is used to get the gate-level netlist of the design in the target technology (standard cells for an ASIC production or FPGAs for prototyping purposes). After a manual partitioning of the design to multiple FPGAs (see Figure 7) and placement and routing of each FPGA, the generated bitstream is downloaded immediately to the FPGA based printed circuit board which is currently under development for this purpose. After a few hours (varying with the chosen generic parameters of the design) the prototype of the information processing part will be ready for testing in its real process environment.

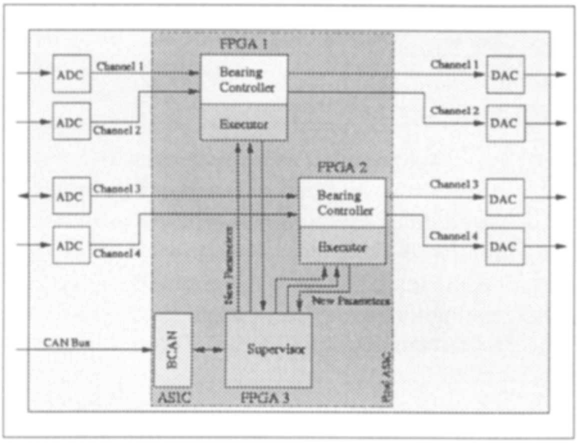

FPGA 1
ADC
Channel 1
Bearing Controller
Executor
Channel 1
DAC
ADC
Channel 2
Channel 2
DAC
FPGA 2
ADC
Channel 3
Bearing Controller
Channel 3
DAC
ADC
Channel 4
Executor
Channel 4
DAC
New Parameters
New Parameters
CAN Bus
BCAN
Supervisor
ASIC
FPGA 3
Final ASIC

Figure 7: Abstract layout of the FPGA based printed circuit board

Because of the unknown complexity of the various control algorithms to follow, only 2 channels are controlled by one FPGA, containing the dataflow oriented bearing controller itself, and an additional controlflow oriented finite state machine called "Executor" receiving new parameters by the master controller, called "Supervisor". After reset, the "Supervisor" initializes the parameters of the bearing controllers and the registers of the BCAN-ASIC, a CAN-Bus Controller, developed previously with ExpressV-HDL as a hierarchical finite state machine and currently in production. The BCAN-ASIC realizes the connection of the magnetic bearing system to an external PC which monitors its operation, allowing a changing of parameters and sensing the state of the bearing system. In an initial design the control algorithm implements a fully parallel PID controller. FPGA 1 and FPGA 2 (see Figure 7) are both programmed with the same design consisting of the bearing control algorithm and the "Executor". The generated code for FPGAs 1 and 2 consist of 640 lines VHDL code and 612 lines for FPGA 3 (the "Supervisor"). Mapping these designs on Xilinx 4013 FPGAs we get the following results:

	No. used	Max. Available	% used
Occupied CLBs	413	576	71%
Bounded I/O Pins:	50	192	26%
CLB Flip Flops:	58	1152	5%

Table 1: FPGA 1 & 2 (“Executor“)

	No. used	Max. Available	% used
Occupied CLBs	182	576	31%
Bounded I/O Pins:	67	192	34%
CLB Flip Flops:	100	1152	8%

Table 2: FPGA 3 (“Supervisor“)

This prototyping environment will be used in the next months to determine the optimal control algorithm for the magnetic bearing system.

After evaluation of different algorithms within the real process environment the final step will be the production of an ASIC containing

- all 4 sensing channels,
- the “Executor", which will be adapted to 4 channels by changing one parameter,
- the “Supervisor" for initialization of the components and
- the BCAN controller for the communication with the outside world.

7. CONCLUSIONS

In this paper we have presented an approach for designing the information processing part of integrated magnetic bearing system by capturing the specification with a high-level graphical environment, automatically generating VHDL code of both the dataflow oriented part as well as the controlflow oriented part of the magnetic bearing controller. To rapid-prototype the system the functionality of the control system is transferred to a DSP system as a fast software prototype or it is mapped as a hardware prototype on several FPGAs on a printed circuit board. This methodology is used to evaluate different control algorithms for the magnetic bearing with respect to the quality of the controller, throughput of the information processing part, time delay needed for the analog/digital and digital/analog converters and area needed for silicon production. The final realization of the information processing part will be an ASIC containing the different blocks described in this paper. The last step includes the verification of the magnetic bearing system consisting only of the magnetic bearing itself and the power amplifier including the ASIC controller.

REFERENCES

1. Haddad, R.A. and Parsons, T.W.: Digital Signal Processing: Theory, Applications and Hardware, Computer Science Press, 1991.
2. Lim, J.S. and Oppenheim, A.V.: Advanced Topics in Signal Processing, Prentice Hall, 1988.
3. Laier, D. and Markert, R.: Simulation of Nonlinear Effects on Magnetically Suspended Rotors. Proc. Of the 1st Int. Conf. On Engineering Computation and Computer Simulation ECCS-1, Changsha, China, 1995.
4. Hoffmann, K.-J. and Markert, R.: Linearisierung von magnetischen Lagerungen für elastische Rotoren per Software. Mechatronik im Maschinen- und Fahrzeugbau, VDI-Berischte 1315, 1997.
5. Le, T. and Renner, F.-M. and Glesner, M.: Hardware-in-the-loop Simulation - A Rapid Prototyping Approach for Designing Mechatronics Systems, IEEE Int. Workshop on Rapid System Prototyping, Chapel Hill, USA, 1997.
6. Glesner, M. and Le, T. and Doan, M.-D. and Kirschbaum, A. and Renner, F.-M.: On the Methodology and Design of Application Specific Processors for Mechatronic Systems. Proc. of the 4th Int. Workshop on Mixed Design of Integrated Circuits and Systems, Poland, 1997.
7. Chang, W.-T. and Kalavade, A. and Lee, E. A.: Effective Heterogeneous Design and Cosimulation. NATO ASI Series, Vol. 310, chapter in Hardware/Software Co-design, Kluwer Academic Publishers, 1996.
8. Sriram, S.: Minimizing Communication and Synchronization Overhead in Multiprocessors for Digital Signal Processing. PhD thesis, University of California at Berkeley, USA, 1995.
9. Lee, E. A. and Messerschmidt, D. G.: Synchronous Data Flow. Proc. of the IEEE, 1987.
10. Drusinsky, D. and Harel, D.: Using Statecharts for Hardware Description and Synthesis. IEEE Trans. on Computer Aided Design, 1989.
11. Kirschbaum, A. and Renner, F.-M. and Wilmes, A. and Glesner, M.: Rapid-Prototyping of a CAN-Bus Controller: A Case Study. Proc. of the 7th IEEE Int. Workshop on Rapid System Prototyping (RSP), Greece, 1996.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronic design of digital electrovalves modulated in PWM

M. Sorli[a] and G. Figliolini[b]

[a] Department of Mechanics, Politecnico di Torino
Corso Duca Degli Abruzzi 24, 10129 Torino, Italy

[b] Department of Industrial Engineering, University of Cassino
G. Di Biasio 43, 03043 Cassino (Fr), Italy

The aim of the actual paper is the evaluation of several parameters that can affect the dynamic and static behavior of PWM modulated digital valves, in order to define some design procedures. Then a specific 2/2 digital valve suitable for PWM applications has been studied. The on/off delay times are evaluated by measuring both input PWM voltage and the current adsorbed by valve. The effects of supply and vent pressures are shown and a preliminary mathematical model is formulated.

NOMENCLATURE

a	Initial width of gap 1	[m]
A_a	Mean area of armaure	[m^2]
A_c	Mean area of core	[m^2]
A_1	Area of gap 1	[m^2]
A_2	Area of gap 2	[m^2]
b	Initial width of gap 2	[m]
i	Coils current	[A]
I_G	Plunger inertia moment	[kgm^2]
K_T	Torsional stiffness	[Nm/rad]
L	Inductance	[h]
l_a	Length of armature flux path	[m]
l_c	Length of core flux path	[m]
l_G	Position of plunger center mass	[m]
l_1, l_2	Mean position of gaps 1,2	[m]
m	Plunger mass	[kg]
N	Number of coil turns	
R	Coil resistance	[Ω]
$\mathcal{R}_1$	Reluctance of gap 1	[h^{-1}]
$\mathcal{R}_2$	Reluctance of gap 2	[h^{-1}]
$\mathcal{R}_a$	Armature reluctance	[h^{-1}]
$\mathcal{R}_c$	Core reluctance	[h^{-1}]
T_d	Driving torque	[Nm]
T_p	Pressure torque	[Nm]
VA	Input voltage	[V]
α	Corrective parameter	
μ_0	Air permeability	[h/m]
μ_r	Relative permeability	[h/m]
θ	Plunger angular displacement	[rad]

1. INTRODUCTION

An automatic system pneumatically actuated can be controlled in analogue environment by using proportional valves in pressure or flow. When the analogue signal control is modulated good performances by using digital valves can be obtained. When a digital 2/2 valve is opened and closed by the fast movement of a small plunger, a variable and quite continuous mass flow-rate can be obtained. This activation can be carried out via hardware by means of a suitable modulation card or via software by using a PLC controller directly. In particular a PWM (Pulse Width Modulation) technique can be used to have a good control by means of low-cost and easy-running digital valves. This solution can be used in the robotics field to control the position/speed of robotic arms or to control the grasp force of grippers. In previous papers [1-3] the grasp force of a gripper pneumatically actuated has been controlled in open loop by using two 2/2 digital valves modulated in PWM. These valves have been mounted in supply and discharge respectively to obtain an equivalent 3/way valve. The control of the grasp force has been obtained by controlling the pressure in the push chamber of the pneumatic actuator by means of a suitable control of the mass flow-rate across the equivalent valve. A theoretical model has been formulated to develop computer simulations and a good agreement with the experimental results has been obtained. A closed loop force control of a double-acting pneumatic actuator has been developed by using four 2/2 digital valves modulated through a combined PNM-PWM (Pulse Number Modulation–Pulse Width Modulation) technique [4-5]. Each pair of valves has been mounted in supply and discharge to each chamber of the actuator. Each valve incorporates nine plungers and each one is controlled by a different electromagnetic. The combined PNM-PWM modulation technique has allowed to control eight plungers through on/off signals and the last one in PWM. A study describing the model of instantaneous mass flow rate across a pneumatic PWM solenoid valve and defining the effect of maximum modulation ratio has been presented in [6]. A pneumatic diaphragm motor controlled by two digital valves modulated in PWM has been studied in [7-8] to substitute the conventional pneumatic servomechanism. The static and dynamic performances have been analyzed by a theoretical and experimental view point. When a PWM modulation is adopted the effects of on/off delay times have been studied to verify the experimental behavior of 2/2 Smc XT311-3 valves.

The aim of actual paper is the evaluation of several parameters that can affect the dynamic and static behavior of PWM modulated digital valves in order to define some design procedures. Then a specific 2/2 digital valve suitable for PWM applications, type Matrix 821.100C224, has been studied. The on/off delay times are evaluated by measuring both input PWM voltage and the current adsorbed by valve. The effects of supply and vent pressures are shown and a preliminary mathematical model is formulated. The good agreement between the numerical and experimental results has allowed to validate the proposed model.

2. FEATURES OF THE VALVE

The 2/2 digital valve Matrix 821.100C224 of Fig.1 is a low-cost and easy-running component which presents a simple structure with small size and weight. A little plunger can rotate about point O by determining the closing or the opening of the valve. The closing movement is obtained through the drop between upstream and downstream pressures, while the opening movement is driven by 24V DC supply signal of the solenoid. The high dynamic performances in terms of on/off delay times makes this type of valve suitable for PWM modulation. The dynamic and static behaviors of the valve have been analyzed by means of a suitable test-bed, Fig.2, mainly composed by a variable voltage supply 1 to impose a Vrif signal control ranging between 0÷10V, by a PWM modulation card 2 driving a solenoid valve 3, by two pressure transducers 4 measuring upstream p_m and vent p_V, by a flow-regulator 5 to vary p_V and finally by a flow meter 6 (ROTA H1.630-880) to measure the flow-rate Q. In order to ensure a low temperature of the electromagnetic circuit of electrovalve a suitable period T=15.6 ms (f=64.1 Hz) has been imposed.

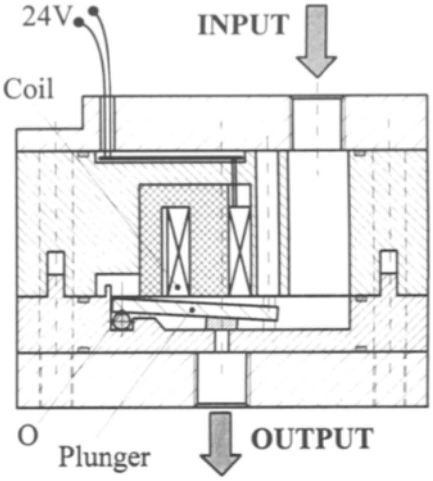


Fig.1 – Digital electrovalve Matrix 821.100C224.

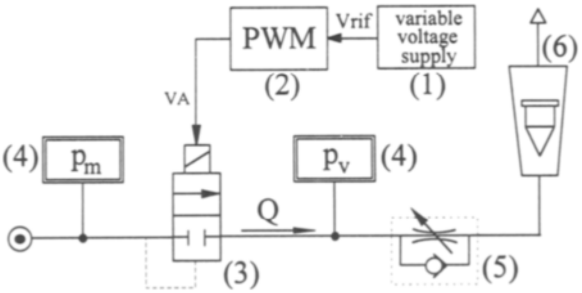


Fig.2 – Experimental set-up.

A first series of tests has been carried out in order to measure the input voltage VA and the current i adsorbed by the valve, maintaining constant p_m=7 bar rel, varying p_v and duty/cycle ratio t/T. Figs.3,4,5 have been obtained for free discharge p_v=0 bar rel, respectively for t/T=20%, 40%, 70%, which correspond to Vrif=2,4,7 V. Analogously in Figs.6,7,8 and Figs.9,10,11 the experimental results with p_v =2 bar rel and p_v=4 bar rel respectively are shown for the previous duty/cycle ratios.

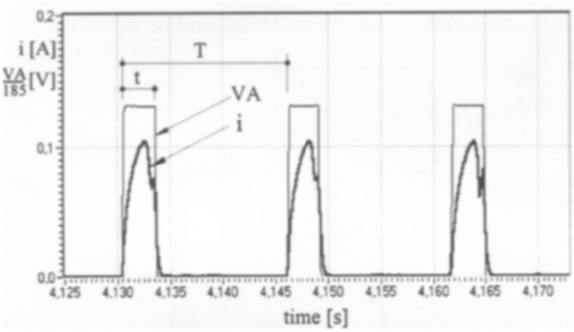


Fig.3 – Input voltage VA and adsorbed current i for Duty/cycle=20%, p_m=7 bar rel and p_v=0.

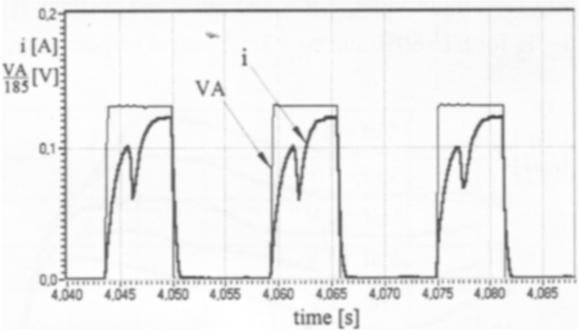


Fig.4 – Input voltage VA and adsorbed current i for Duty/cycle=40%, p_m=7 bar rel and p_v=0.

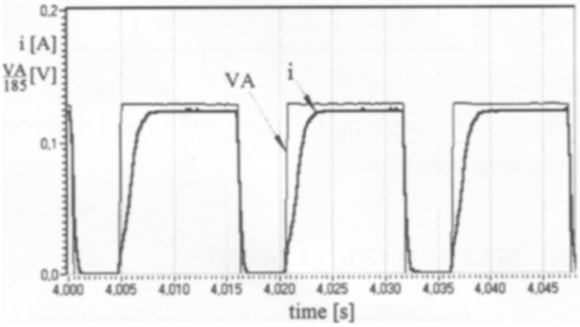


Fig.5 – Input voltage VA and adsorbed current i for Duty/cycle=70%, p_m=7 bar rel and p_v=0.

The current diagrams have been obtained by measuring the potential difference across a shunt resistor which has been installed in series with the valve coil. When an input voltage is given the current begins to circulate across the electric circuit to reach the steady-state value. During this transient the current diagram presents a typical peak in correspondence of opening start of the plunger. For all vent pressure conditions this behavior occurs until t/T is approximately equal to 70%.

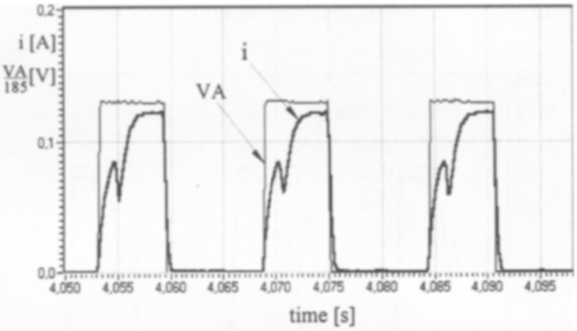


Fig.6 – Input voltage VA and adsorbed current i for Duty/cycle=20%, p_m=7 bar rel and p_v=2 bar rel.

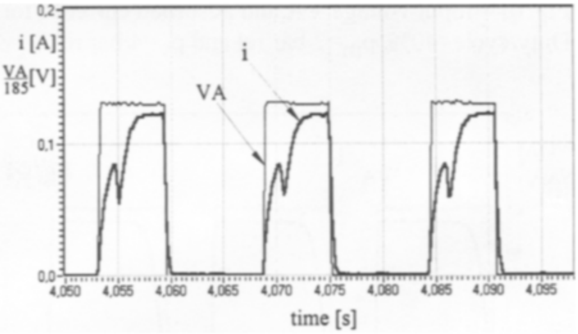


Fig.7 – Input voltage VA and adsorbed current i for Duty/cycle=40%, p_m=7 bar rel and p_v=2 bar rel.

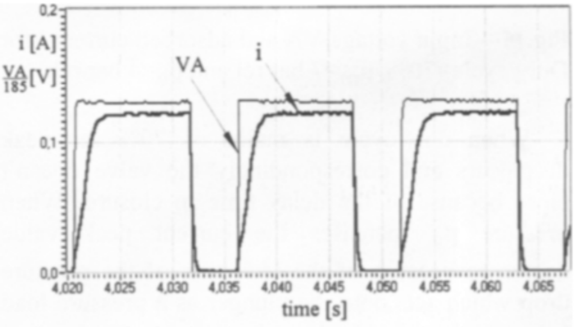


Fig.8 – Input voltage VA and adsorbed current i for Duty/cycle=70%, p_m=7 bar rel and p_v=2 bar rel.

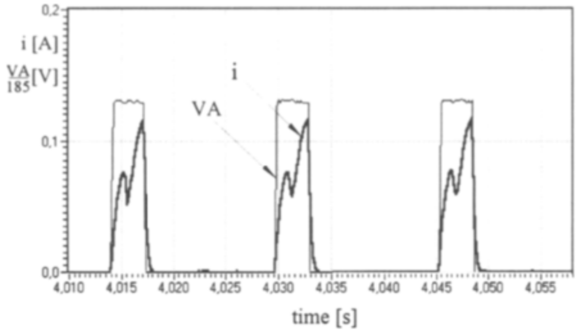


Fig.9 – Input voltage VA and adsorbed current i for Duty/cycle=20%, p_m=7 bar rel and p_v=4 bar rel.

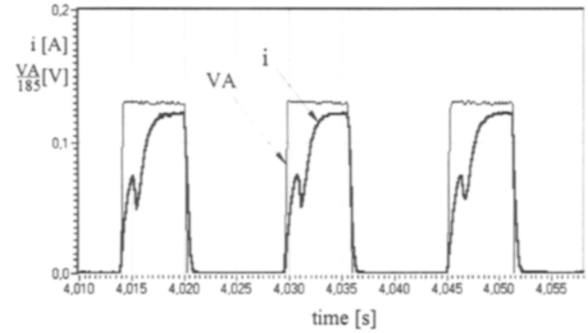


Fig.10 – Input voltage VA and adsorbed current i for Duty/cycle=40%, p_m=7 bar rel and p_v=4 bar rel.

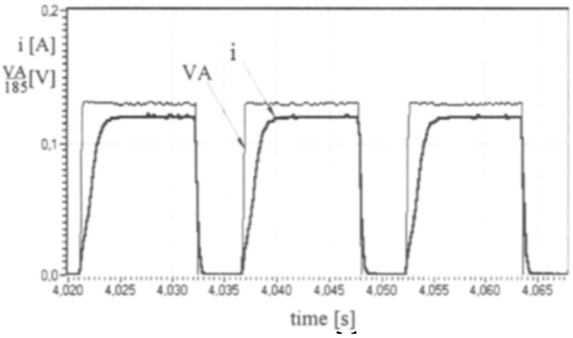


Fig.11 – Input voltage VA and adsorbed current i for Duty/cycle=70%, p_m=7 bar rel and p_v=4 bar rel.

When duty/cycle is major of 70% the peak disappears and correspondingly the valve doesn't close because of the delay time in closure. When pressure p_v increases the current peak value decreases because of the reduction of the pressure drop which acts onto the plunger as a pressure load torque at the opening time. A second tests campaign has been devoted to the evaluation of the influence of upstream pressure p_m.

From Fig.12 can be noted as varying p_m from 3 to 5 to 7 bar rel with a fixed duty/cycle=60% and free discharge p_v=0, the current peak value increases as indicated by i_1, i_2 and i_3 diagrams respectively.

Other experimental tests have been carried out to determine the flow-rate Q [Nl/min] across the digital valve as function of duty/cycle. Maintaining p_m equals to 7 bar rel, different flow-rate curves versus p_v for several Vrif reference signals or duty/cycles of the PWM modulation, are shown in Fig.13. When pressure drop is smaller than 1 bar rel, the plunger remains opened and the modulation has not effect.

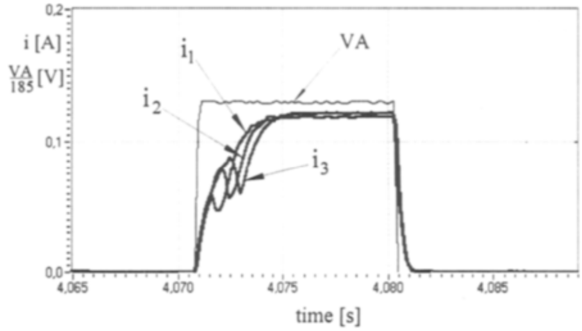


Fig.12 – Input voltage VA and adsorbed currents i_1, i_2, i_3 for t/T=60% and p_m=3;5;7bar rel respectively.

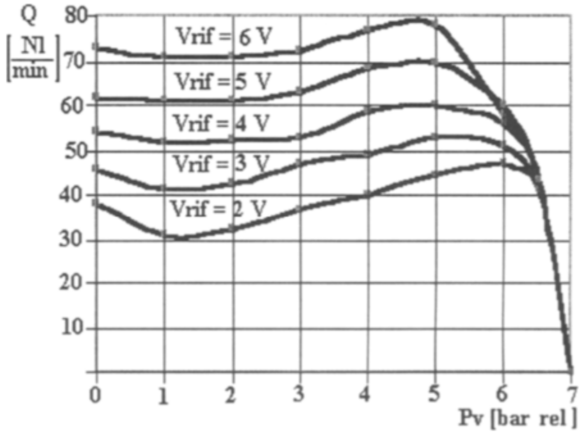


Fig.13 – Q versus p_v for p_m=7 bar rel and different Vrif reference signals.

3. MATHEMATICAL MODEL

A first model of the valve has been formulated in order to evaluate the dynamic variation of the adsorbed current by coil, for several conditions of duty/cycle. The aim is the characterization of the possible opening conditions of the plunger. The lay-out of Fig.14 has been assumed.

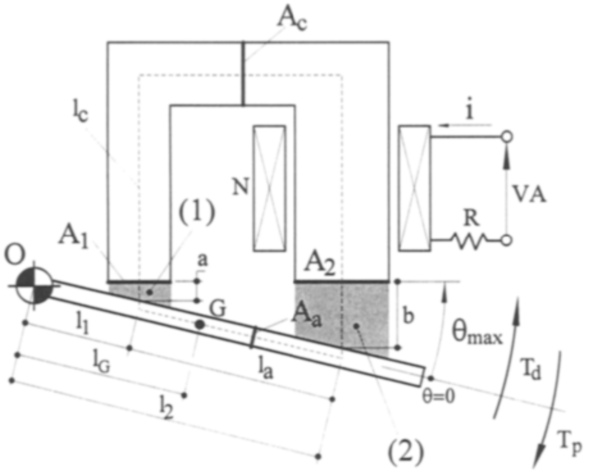


Fig.14 – Schematic lay-out of the valve.

The reluctances $\Re_1$, $\Re_2$ of air gaps 1 and 2 are function of angle θ which defines the plunger angular displacement, while the reluctances $\Re_a$, $\Re_c$ of plunger and core are considered constants. They can be expressed respectively as

$$\Re_1 = \frac{a - l_1 \vartheta}{\mu_0 A_1} \tag{1}$$

$$\Re_2 = \frac{a - l_2 \vartheta}{\mu_0 A_2} \tag{2}$$

$$\Re_a = \frac{l_a}{\mu_0 \mu_r A_a} \tag{3}$$

$$\Re_c = \frac{l_c}{\mu_0 \mu_r A_c} \tag{4}$$

Thus the total inductance L of the circuit becomes

$$L = \frac{N^2}{\Re_1 + \Re_2 + \Re_a + \Re_c} \tag{5}$$

The voltage equation of the solenoid can be expressed as

$$VA = Ri + L\frac{di}{dt} + i\frac{dL}{dt} \tag{6}$$

The driving torque T_d acting on the plunger can be given by

$$T_d = \frac{1}{2} i^2 \frac{dL}{d\vartheta} \alpha \tag{7}$$

where α is a corrective parameter which is assumed in simulation tasks. The dynamic equilibrium of plunger about joint O for $0 \le \theta \le \theta_{max}$ gives

$$T_d = \left(I_G + m l_G^2\right)\frac{d^2\vartheta}{dt^2} + K_T \vartheta + T_p \tag{8}$$

where K_T is a torsional stiffness of joint O.

The system of equations (1)÷(8) has been numerically integrated with a suitable model written in Matlab-Simulink environment and several simulations have been carried out by neglecting the effect of pressure torque Tp. This effect has been compensated in first approximation by using the corrective parameter α.

A good agreement with the experimental behavior of the valve at p_m=7 bar rel and p_v=0 bar rel can be observed by Figs.15-17 in which data of Tab.1 have been used.

Table 1 – Input data

$a = 0.13$ [mm]	$l_G = 4.5$ [mm]
$A_a = 1*10^{-6}$ [m^2]	$l_1 = 2.6$ [mm]
$A_c = 4.8*10^{-6}$ [m^2]	$l_2 = 8$ [mm]
$A_1 = 11.2*10^{-6}$ [m^2]	$m = 0.7275*10^{-3}$ [kg]
$A_2 = 19.6*10^{-6}$ [m^2]	$N = 1725$
$b = 0.4$ [mm]	$R = 200\ \Omega$
$I_G = 4.9*10^{-9}$ [kgm^2]	$\alpha = 0.37$
$K_T = 2*10^{-3}$ [Nm/rad]	$\mu_0=4\pi*10^{-7}$ [h/m]
$l_a = 5.4$ [mm]	μ_r=700 [h/m]
$l_c = 30$ [mm]	$\theta_{max} = 0.05$ [rad]

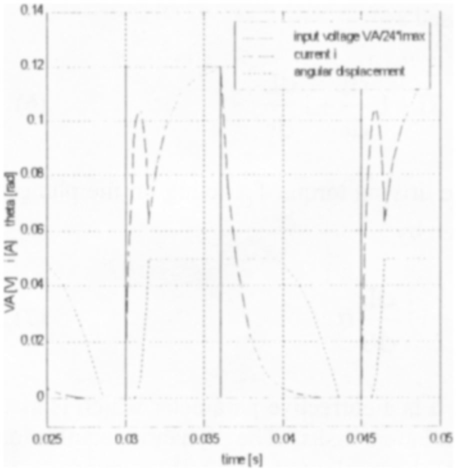


Fig.15 – Duty/cycle=40%, pm=7 bar rel, pv=0

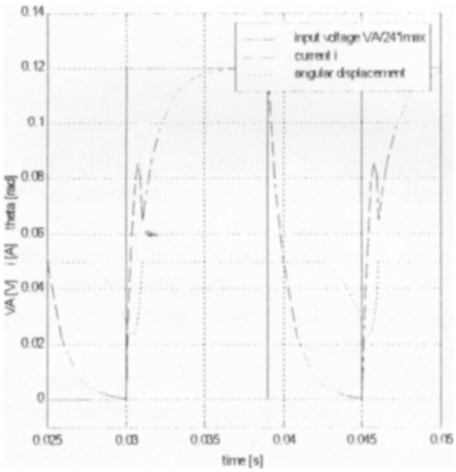


Fig.16 – Duty/cycle=60%, pm=7 bar rel, pv=0

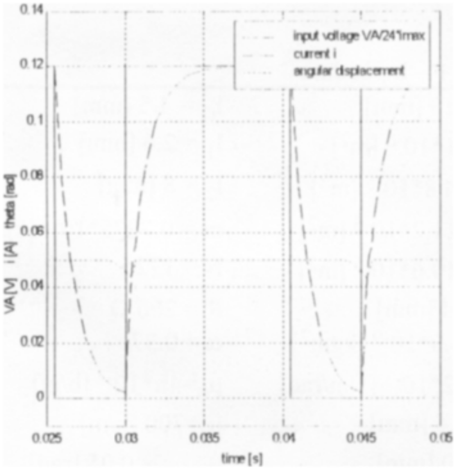


Fig.17 – Duty/cycle=70%, pm=7 bar rel, pv=0

4. CONCLUSION

On the basis of the actual results presented in the paper, the research will be addressed to estimate the effect of the modulation carrier frequency on the plunger displacement measured via a laser sensor and to consider in details the pressure load disturbances.

REFERENCES

1. Figliolini G., Gastaldi L., Sorli M., Analisi teorica e sperimentale di un sistema di controllo della pressione in anello aperto con valvole digitali modulate in PWM, XII National Congress AIMETA, Vol.III, (1995) pp.99-104, (in italian)
2. Figliolini G., Sorli M., A pressure control system with digital electrovalves modulated in PWM and controlled via PLC, 6th Int. Workshop on Robotics in Alpe-Adria-Danube-Region RAAD, Cassino, (1997) pp.453-458.
3. Figliolini G., Sorli M., Controllo forza di gripper ad azionamento pneumatico, XIII National Congress AIMETA, Siena, Vol. II, (1997) pp.249-254, (in italian)
4. Sorli M., Ferraresi C., Pastorelli S., Force controlling pneumatic servoactuator via digital PWM modulated valves, 2nd Int. Symp. on Fluid Power Trans. and Control ISFP, Shangai, (1995) pp.480-485.
5. Sorli M., Ferraresi C., Pastorelli S., Development of a new linear electro-pneumatic servoactuator, Int. Symp. on Ind. Robots ISIR, Milano, (1996) pp.431-436.
6. Ye N., Scavarda S., Betemps M., Jutard A., Models of a pneumatic PWM solenoid valve for engineering applications, ASME Journ. Of Dyn. Syst. Measur. And Control, Vol.114, Dec.1992, pp.680-688.
7. Yang C., Kawaguchi H., Kawai S., Kawakami Y., Machiyama T., Some considerations on improvement of stiffness for pneumatic diaphragm motors, 2nd Int. Symp. on Fluid Power Trans. and Control, Shanghai, (1995) pp.425-430.
8. Yang C.H., Kawai S., Kawakami Y., Sinozaki K., Machiyama T., Experimental study on solenoid valves controlled pneumatic diaphragm motor, 3rd Int. Symp. on Fluid Power Trans. and Control, (1996) pp.247-252.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Computer Aided Conceptual Design Method for Mechatronic Systems in Process Oriented Heavy Machinery

St. Dierneder [a], R. Scheidl [a], K. Mörwald [b] and J. Guttenbrunner [b]

[a] University of Linz, Department for Foundations of Machine Design
Altenbergerstrasse 69, A-4040 Linz, Austria

[b] VÖEST-Alpine Industrieanlagenbau GmbH
Turmstrasse 44, P.O. Box 4, A-4031 Linz, Austria

Process oriented heavy machinery requires a clearly structured, functional oriented design method. Mechatronization is seen as a means to achieve that goal particularly under the condition of an extended automation. For the conceptual design of such systems *Functional Decomposition* is a well fitted method which aims at a hierarchical structure of functions and sub systems, which are technically realised by mechatronic modules. Computer Assistance should advise the design team to do its work in a systematic manner, to get an automatic documentation, to help finding possible technical realizations, and to assist further detail design. The implementation of these concepts in high level rather standard software and the handling of the depiction method on the computer is reported.

1. INTRODUCTION

We currently are particularly interested in process oriented heavy machinery e. g. metallurgical plant equipment. Process oriented heavy machinery is characterised by the following issues

- Customer specific design (at least in parts)
- Complex process technology
- High level automation, because the costs for automation devices are relatively small related to the costs of the often heavy mechanical components and, if it improves performance, it mostly pays within short time
- Need for thorough engineering, because corrections of defective components are expensive and a testing under true operational conditions is mostly impossible
- A very cautious clientele with respect to innovations, because a malfunctioning machine or plant can be economically disastrous.

Nevertheless, the strong market and technology dynamics of today demands for shorter innovation cycles and reduced development costs of new products or new product generations. Short cycle innovation not only requires to shorten the development process but also, to speed up and intensify the marketing process of a new product or technology. The stronger and stronger „mechatronization“ of many products has impact also on computer aided design and on design methodologies.

2. MECHATRONIZATION OF PROCESS MACHINERY

These characteristics of process oriented heavy machinery, as mentioned above, require specific issues in mechatronic design. Concepts, which are successful in other branches of mechanical engineering like in precision engineering, where mechatronization was realised first, cannot be transferred straight forward to this field.

By „mechatronization“ we mean – following a definition by J. Buur [1] – the increase of functional interaction and spatial integration of components from different engineering fields.

Functional interaction is reality in metallurgical machinery, where automation started already 25 years ago and was extended qualitatively and quantitatively ever since. This automation included improved or new sensing, automation and communication devices, resulting in an extended automation network with a multiply levelled hierarchical structure, application of new and advanced control concepts, and, finally, advanced actuation. But spatial integration often did not take place for good reasons, e.g., the often very adversarial environmental conditions, the need to do fast maintenance and therefore to have good access to such devices. To gain benefits from the spatial integration specific design maxims have to be developed, which have to meet the following essential requirements

- Achieve high reliability and robustness of the technical system under the harsh environmental conditions
- Promote modularization, that is the split up of machines into well defined rather autonomous functional units
- Provide a sustainable and easy accessible documentation reporting the development of a certain component, the practical experiences, and the reasons for the selection of a specific solution
- Standardise on the one hand but provide high flexibility to adopt to specific customer requirements on the other hand

Which benefits offers spatial integration for heavy machinery? In particular, which of these benefits are relevant for the buyer of such machines?

We believe that modularization and methodological mechatronic design of machines are important roots to a beneficial mechatronization in heavy machinery. The modules should be mechatronic systems themselves and should have a minimum coupling to the rest of the machine, thus they are less or more independent units devoted to a certain function.

This approach has several advantages

- It avoids unnecessary complex systems with a lot of couplings. In general, strongly coupled systems are difficult to understand and to overview, and hence are difficult to be modified reliably and also cause troubles in maintenance
- It improves reliable and fast maintenance, because these units (modules) can be replaced at once by a spare unit, and maintenance work then can take place in an appropriate workshop with the required surrounding conditions
- It provides a save route to innovation by small steps with limited risk
- It reduces development time and costs as well as of production costs
- It increases functionality and quality
- It supports innovative ideas in all phases of the design process of new developed products as well as for upgrading existing products because of enforcing a clear orientation on functions first rather than on technical solutions
- It contributes to reliability despite the fact that sensitive components may come close to areas with very harsh conditions

The last advantage can only become reality if new solutions for placing electronic components reliably into adversarial environmental conditions are available. Similar conditions exist in car engineering, where a lot of electronic components nowadays are installed. The big difference between cars and heavy machinery cannot be denied, but the successful mastering of initial problems with machine near electronics in car engineering makes us confident that this can be done in process oriented machinery too.

Spatial integration offers several opportunities in this respect.

- Use stable mechanical components as housing for the electronics to prevent mechanical, thermal, and electromagnetic harm
- Reduce cabling and the number of connectors, which often are sources of failure and disturbances

- Reduce complexity and thus contribute to reliability by an appropriate mechatronic design which also should aim at a simple solution considering all the different components.

The importance of a conceptual design strategy for mechatronization is twofold:

- Help to find basic concepts as well as technical solution principles for this mechatronization approach
- Supply rules to find optimal solutions, when these general concepts are adopted to the specific design of an individual project

3. THE FUNCTIONAL DECOMPOSITION METHOD

The methodological basis of our approach is a clear *Functional Decomposition* of the overall system. We call this the *"System Under Consideration" (SUC)*. The *Functional Decomposition* proceeds top down and creates a tree like structure (a graph) involving both the functions required from the overall and all sub systems as well as the technical solutions for a certain function. For the further explanation we refer to Figure 1 and Figure 2.

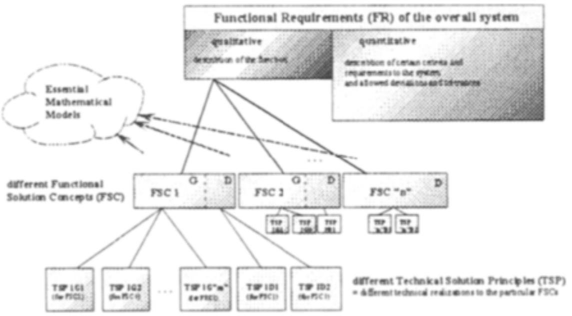

Functional Requirements (FR) of the overall system
qualitative
definition of the function
quantitative
description of certain criteria and requirements to the system and allowed deviations and tolerances
Essential Mathematical Models
different Functional Solution Concepts (FSC)
FSC 1
G
D
FSC 2
G
D
FSC "n"
D
TSP 1G1 (for FSC1)
TSP 1G2 (for FSC1)
TSP 1G"m" (for FSC1)
TSP 1D1 (for FSC1)
TSP 1D2 (for FSC1)
different Technical Solution Principles (TSP) = different technical realizations to the particular FSCs

Figure 1. Schematic presentation of the functional decomposition of the overall system – the system under consideration (SUC).

At the beginning of the conceptual design phase either because of a fully new design or a redesign the *Functional Requirements* (FRs) should be defined in a clear an terse way. We follow here Nam P. Suh's 2nd Design Corollary "Minimize the number of FRs and constraints" (see [2] page 52). This is a very important and challenging task and the success of a systematic approach to design depends strongly on this first step. The definition of these requirements comprises a precise qualitative description of the desired functionality and quantitative measures of the essential performance data e. g. tolerances. For the further explanation of the handling of the FRs including the definition of so called *"Essential Mathematical Models* (EMMs)", "Active Functions (AFs)", and "Passive Functions (PFs)" we refer to [4].

This methodological design consists of three main steps.

1. Finding and defining the FRs for the system under consideration (SUC)
2. Finding solutions, which perform the required functions. – For a particular function different technical solutions can exist, each of them could be depicted in a so called Functional Solution Area (FSA).
3. Evaluating these solutions with respect to the Functional Requirements to choose the best.

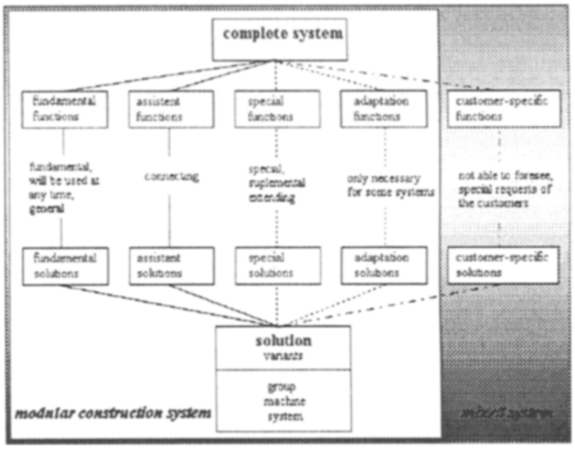

complete system
fundamental functions
assistent functions
special functions
adaptation functions
customer-specific functions
fundamental, will be used at any time, general
connecting
special, supplemental extending
only necessary for some systems
not able to foresee, special requests of the customers
fundamental solutions
assistant solutions
special solutions
adaptation solutions
customer-specific solutions
solution
variants
group machine system
modular construction system

Figure 2. The result of the Functional Decomposition (see [3]).

4. COMPUTER ASSISTANCE

As pointed out in the introduction, computer aided conceptual design in the engineering branch under consideration is not limited to an early phase of the product life cycle. When a system (overall system, sub-system, component) is designed totally new the conceptual design normally is a mentally intensive and challenging work. Therefore, it is important for the success of such design concepts, that they are given in the hand of excellent engineers.

Computer Assistance in the phase of conceptual design should aim at the following issues

- seduce the engineer to do conceptual design in a systematic manner
- provide support for the documentation of the design work
- release the design team from routine work
- give good support for the later detail design work
- support standardisation

The programming environment uses exclusively high level and widespread software. The kernel is realised by MATLAB/SIMULINK®. Its block oriented description of dynamic systems is used for the representation of the structure of the system both with respect to the *Functional Decomposition* and the technical solutions. New functional parts can be easily defined and arranged in appropriate or new libraries. The *Functional Decomposition* is documented in MATLAB/SIMULINK. The graphical interface is shown in Figure 3. SIMULINK provides an easy way to handle block oriented structures. A block itself can be another complex structure which is represented in a special window and can be masked and unmasked. This fits well to the hierarchical structure of *Functional Decomposition.* SIMULINK also offers a well defined and easy to handle interface to other software programs, like AUTOCAD®, MAPLE V® and so on.

Besides blocks for the definition of functions, library blocks with established technical solutions to frequently required functions are provided. They are also classed in a hierarchical structure from the general down to the more specific.

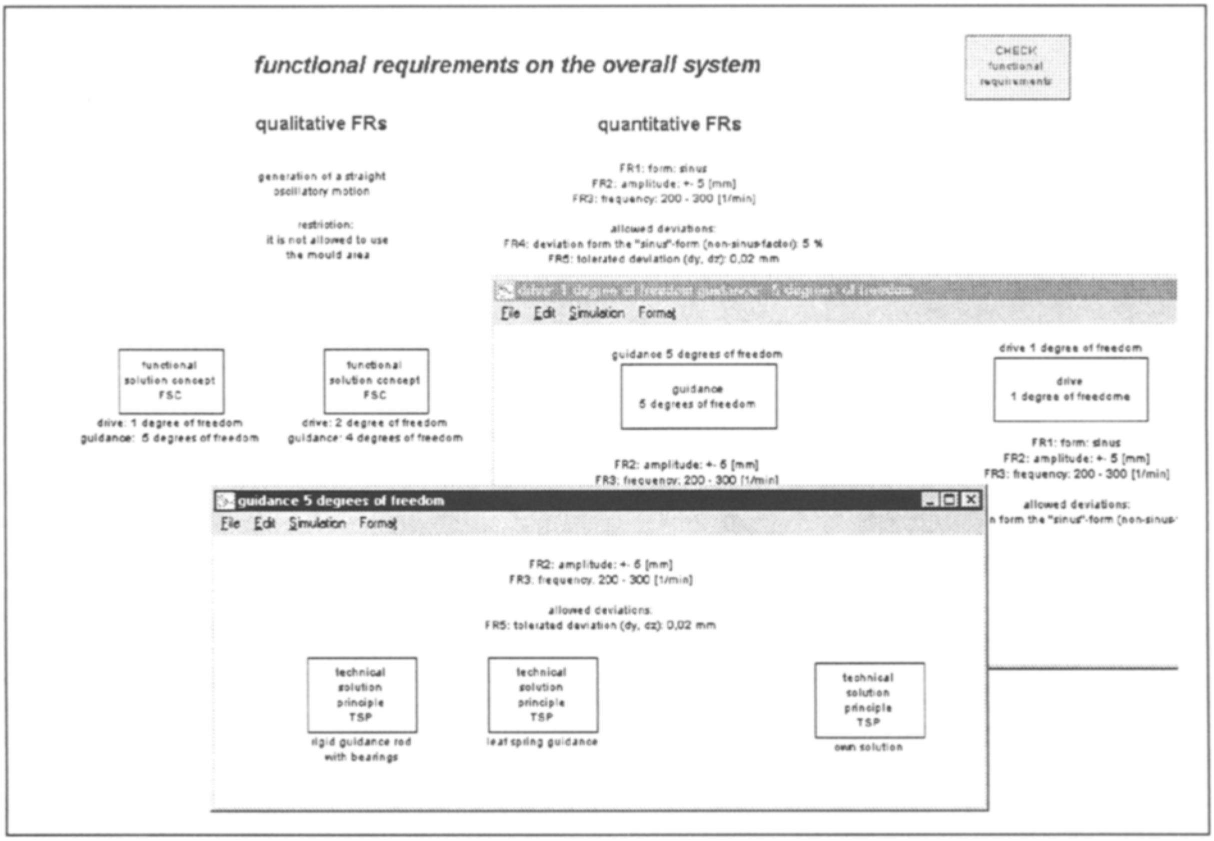

functional requirements on the overall system
CHECK functional requirements
qualitative FRs
quantitative FRs
generation of a straight oscillatory motion
restriction: it is not allowed to use the mould area
FR1: form: sinus
FR2: amplitude: +- 5 [mm]
FR3: frequency: 200 - 300 [1/min]
allowed deviations:
FR4: deviation form the "sinus"-form (non-sinus-factor): 5 %
FR5: tolerated deviation (dy, dz): 0,02 mm
File Edit Simulation Format
functional solution concept FSC
drive: 1 degree of freedom
guidance: 5 degrees of freedom
functional solution concept FSC
drive: 2 degree of freedom
guidance: 4 degrees of freedom
guidance 5 degrees of freedom
guidance 5 degrees of freedom
drive 1 degree of freedom
drive 1 degree of freedome
FR2: amplitude: +- 6 [mm]
FR3: frequency: 200 - 300 [1/min]
FR1: form: sinus
FR2: amplitude: +- 5 [mm]
FR3: frequency: 200 - 300 [1/min]
allowed deviations:
n form the "sinus"-form (non-sinus-
guidance 5 degrees of freedom
File Edit Simulation Format
FR2: amplitude: +- 6 [mm]
FR3: frequency. 200 - 300 [1/min]
allowed deviations:
FR5: tolerated deviation (dy, dz): 0,02 mm
technical solution principle TSP
rigid guidance rod with bearings
technical solution principle TSP
leaf spring guidance
technical solution principle TSP
own solution

Figure 3. General structure of the Functional Decomposition of an oscillatory-drive.

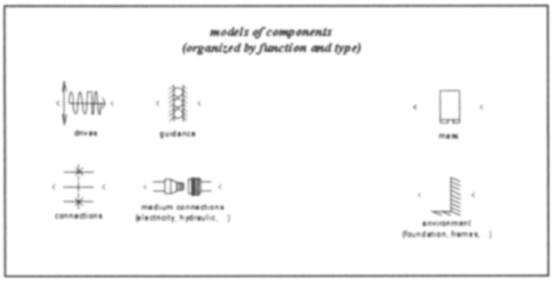

models of components
(organized by function and type)
drives
guidance
mass
connections
medium connections
(electricity, hydraulic, ...)
environment
(foundation, frames, ...)

Figure 4. The existing library box of available components in SIMULINK.

The Figures 4 and 5 give examples of the currently available libraries. Each of the components is related to one or more mathematical models to do dimensioning or some other sort of analysis. These models are implemented with different software, ranging from standard table calculation to more sophisticated dynamic simulation packages.

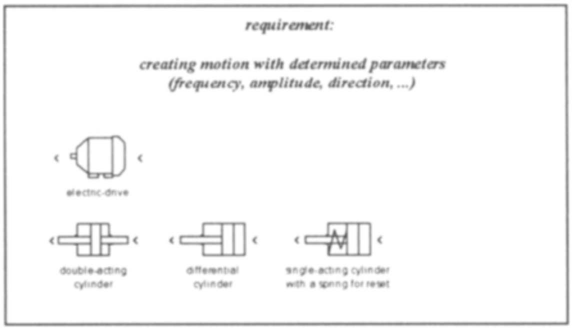

requirement:
creating motion with determined parameters
(frequency, amplitude, direction, ...)
electric-drive
double-acting
cylinder
differential
cylinder
single-acting cylinder
with a spring for reset

Figure 5. An example-library-box for the drive-component from Figure 4.

We explain the handling of the proposed approach by the example – generation of a straight oscillatory motion.

At the upper level the *Functional Requirements* have to be defined as shown in Figure 3. This *Functional Decomposition* and the definition of the functions has to be done for the *System Under Consideration SUC*. After this, one is able to create a structure of a specific solution, which details out a chosen *Technical Solution Principle (TSP)*, using the icons and blocks of the mechanical components, for which examples are given in Figure 4 and 5.

Finally, on lower level the system is represented by blocks and connections as can be seen in Figure 6 and Figure 7, including more and more detailed information with respect to the individual components and parts of the system.

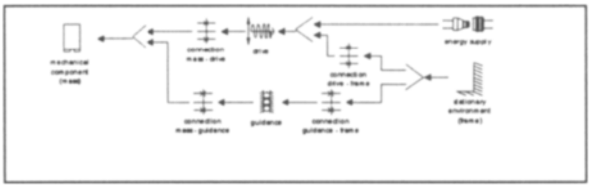

mechanical
component
(mass)
connection
mass - drive
drive
energy supply
connection
drive - frame
stationary
environment
(frame)
connection
mass - guidance
guidance
connection
guidance - frame

Figure 6. General structure of the system (drive, guidance and frame are shown only as blocks).

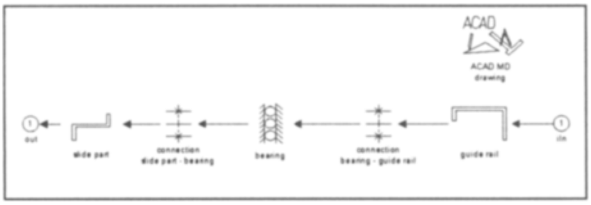

ACAD
ACAD MD
drawing
out
slide part
connection
slide part - bearing
bearing
connection
bearing - guide rail
guide rail
in

Figure 7: Detailed structure of the guidance.

To meet the requirement for standardisation and flexibility the *Parametric Variant Design* is supported. To use this possibility each parametrically adjustable system must be linked to mathematical models for a proper dimensioning and simulation. Therefore, you also have to take care to keep track of the different variants in computation and design during the development cycle. Another means to support reliable design is to make practical experiences with certain technical solutions or features available to designers in a comfortable and stimulating manner. Drawing is done in AUTOCAD® because this will be the preferred system of our industrial partner in the future. The design of new mechanical systems will be done in 3D and, wherever useful, as a parametric design. This provides high flexibility despite preserving a certain standard. In other words, more flexibility is achieved by standardisation at a higher level of abstraction. The drawing files are linked to the structure established in MATLAB®.

There also supervision of the specific parameter values is done to assure that, for instance, computation is done with the same data as the parametric design.

But SIMULINK also has a few unpleasant shortcomings, e. g., the cumbersome creation of graphical symbols to depict the SIMULINK-blocks and sub-systems as shown in Figure 4 and Figure 5, the strict requirement for a very accurate definition of the data- and path-structure on the computer-system which makes far too much trouble for the user, and the need to define remarks and the FRs as strings, because SIMULINK does not support complex data structures and graphical representations for such cases.

By addressing the FRs by names it is possible to check the completeness of the FRs on each layer by performing a short MATLAB-program.

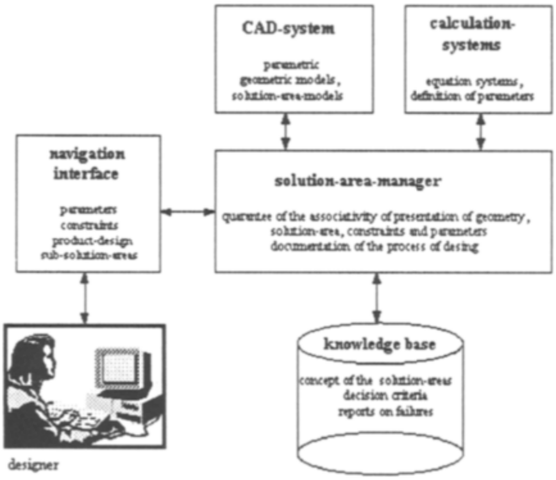

CAD-system
parametric
geometric models,
solution-area-models
calculation-
systems
equation systems,
definition of parameters
navigation
interface
parameters
constraints
product-design
sub-solution-areas
solution-area-manager
quarantee of the associativity of presentation of geometry,
solution-area, constraints and parameters
documentation of the process of desing
knowledge base
concept of the solution-areas
decision criteria
reports on failures
designer

Figure 8: Architecture of this computer aided engineering system.

Finally it must be mentioned, that specific attention was given to the integration of an existing knowledge base filled with practical experiences. This knowledge base includes recommendations on specific design issues, like material selection, reports on failures, arguments and decision criteria for the selection of a certain technical solution. A concept was developed which provides machine part, functional, or date oriented searches as well as the easy and well documented addition of such knowledge. The structure of this system for computer aided engineering is shown in Figure 8. It should be emphasized that in each layer (complete system and detailed structures) of the *Functional Decomposition* (Figures 3 - 7) all special aides like knowledge data, change remarks, calculation files, etc. are available. Thus, it is ensured that the designer gets full support of the system during every stage of the design process.

5. SUMMARY AND OUTLOOK

An essential benefit of mechatronization in process oriented heavy machinery is to get a modular design of the overall system (see Figure 2), composed of mechatronic modules, which correspond widely to the functional structure of the machine. To generate a functional structure appropriately in the conceptual design phase the *Functional Decomposition Method* is recommended. In general, various functional structures can be found to perform the required overall and sub-functions, thus selection steps for the most appropriate one have to be made. This selection should be facilitated by a mathematical model of a certain specific Functional Solution Concept as mentioned in [4] and Computer assistance should guide the design team through the whole conceptual design process, by enforcing some systematic approach.

Further work will concentrate on testing this concept for various applications, as mentioned above, in metallurgical machinery.

REFERENCES:

1. Buur J., *Does mechatronics need a special design attitude?*, Proc. Mechatronics Research Conf., 13th-15th Sept. 1990, St. Albans, England, I MECH E
2. Nam P. Suh, *The Principles of Design,* Oxford University Press, New York, Oxford, 1990
3. Pahl, Beitz, *Konstruktionslehre,* 3. neubearbeitete und erweiterte Auflage, Springer, Berlin – Heidelberg, 1993
4. Scheidl R., Dierneder St., Mörwald K., *Computer Aided Conceptual Design by a Functional Decomposition Method for Process Oriented Heavy Machinery and its Relations to Mechatronization*, Proc. CACD'98, 27th and 28th May, Lancaster

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Computer Based Tool to support Electronics Design in a Mechatronic Environment

R. M. Walters[a], D. A. Bradley[b] and A. P. Dorey[a]

[a]Engineering Department, Lancaster University,
Lancaster LA1 4YR, UK

[b]School of Engineering, University of Abertay; Dundee,
Bell Street, Dundee DD1 1HG, UK

The achievement of a successful mechatronic product or system requires the design team to integrate a range of differing technologies, often with conflicting requirements. Following the initial partitioning of the overall design into its mechanical, electronic and software elements, the remainder of the design process must proceed taking into account the impact of decisions made in one area of technology on the structure and operation of the other technologies involved. This problem is particularly complex in respect of the design of the electronic element of a mechatronic system where no tools comparable to those in either of the mechanical or software domains are available to support and assist the designer during the decision making stages of the design process. The paper therefore sets out an outline for a design support tool for the electronic designer based on the concept of functional modelling which can be linked in to other computer based tools operating in both the mechanical and software domains to support the overall mechatronic design process.

1. INTRODUCTION

The driving force behind the mechatronic concept in design is the ability to transfer complexity from the mechanical domain to the electronics and software domains and has been supported by the continuing increase in capability and reducing real costs of electronic components and in particular of processing power. The resulting technical integration has not however been supported or paralleled by a similar integration of the design methods, methodologies and tools used to support the individual areas of technology.

Historically, each of the major technical disciplines involved in the design of mechatronic systems has been associated with its own individual approach to design. In the case of electronics the driving force behind the development of design tools has been the need to deal with the rapidly increasing complexity of electronics and, in particular the need to support the production of complex circuits on-chip.

Indeed, as the complexity of individual chips increases limitations may well be set by the ability of the designers to comprehend the complexities and applications of the on-chip structures rather than by the electronic hardware.

In design terms, a circuit comprising a few and perhaps tens of components can be reproduced on a single piece of paper and held in the mind of the individual designer. As the number of components in the circuit increases, the ability of the designer to understand the detail of overall structure is correspondingly reduced. Indeed, systems which are themselves formed as the assembly of many thousand components on a single chip may themselves become components of larger designs.

The introduction of computer based tools and simulation enables the total layout to be stored while allowing the designer to concentrate on a particular part or region of the overall design. The computer is also able to ensure that specified design rules are complied with throughout the design, regardless of scale. The combination of rules governing the layout and interconnection of on-chip modules taken together with the available simulation techniques has provided the semiconductor industry with the means of sustaining its growth, producing ever more complex assemblies of components on-chip. Indeed, it has been argued that the semiconductor industry, far from being mature is in a continuous state of rapid evolution! The result is that existing systems are rapidly superseded by newer designs as any observer of the computer industry can readily testify.

In order to support this growth in complexity of electronics, tools such as the Hardware Description languages (HDLs) were developed which allowed an electronic circuit to be described as lines of code. One of this group of languages, VHDL, was adopted as a standard in 1987 and updated in 1993 as a digital hardware description language. The ability to deal with analogue circuits was then added in 1995. The availability and use of a description language such as VHDL as opposed to the use of schematic capture remains a source of debate with the two being used together in practice.

At the moment, HDLs have the advantage of portability, supporting the transfer of designs between the system designer and the fabricator. Despite modularity and the ability to make existing modules available for later use, the complexity of electronics and the rate of change also implies that each design tends to be undertaken from scratch rather than be evolutionary in the manner of many mechanical designs. While electronics designers do have available to them a large number of standard circuit assemblies, the range of possible solutions and the desired variations and problem diversity tends to largely mitigate against an evolutionary strategy [1,2].

2. THE EVOLUTION OF ELECTRONIC DESIGN METHODS AND MECHATRONICS

Prior to the 20th century, engineering design was principally concerned with mechanical and structural systems and placed an emphasis on features such as spatial relationships, motions and the application and transmission of forces. Electrical engineering and, particularly from the middle of the century onwards, electronics evolved along a largely separate path in which the design process placed the emphasis on the manipulation of currents and voltages.

Where electrical engineering, electronics and mechanical engineering did come together it was in the form of a variety of electro-mechanical systems such as the use of relays to control a sequence of essentially mechanical operations.

The advent in 1948 of the transistor and the subsequent growth of the electronics industry led to the development in the 1960's of control mini-computers such as the PDP-8 series. Such computers, though a major advance in the control of plant, remained relatively limited in their scope and capability and it was not until the advent of VLSI technologies and the first 8-bit microprocessors in 1972 that low cost computing power began to impact upon the design and operation of a wide range of products and processes.

The subsequent fall in the real cost of computing power by a factor of the order of 10,000 over the period from 1972 to 1997 has resulted in a massive increase in the deployment of electronic sub-systems across a wide range of products and systems. This development has, as suggested by figure 1, resulted in the concept of mechatronics evolving as an integrated and integrating approach to the design of complex products and systems in which expensive mechanical solutions are being replaced by lower cost solutions based on electronics and software together with simpler mechanisms.

However, this same period has not seen an equivalent coming together of design methodologies to support the integration of electronics, mechanical engineering and information technology represented by the mechatronic concept. Instead, the core disciplines have concentrated on the development of design support tools structured to meet their specific needs.

Despite the introduction of managerial and organisational concepts such as concurrent engineering which attempt in part to address the problem of achieving integration the result remains to often one of an increasing divergence of methods and tools while trying to achieve greater integration across a range of technologies.

3. ELECTRONICS DESIGN AND MECHATRONICS

In developing the 'best' design, Ellis and Ludwig have suggested the following factors as significant:

[The optimal solution may be:]

- That configuration in which the least requirements .... are placed on the [system] units.
- That configuration in which the information is most readily passed from one unit to another.
- That configuration in which the inter-unit equipment .... is kept to a minimum.
- That configuration in which the inter-unit translations .... are kept to a minimum.

The most important measure of any design is however whether or not it meets the defined criteria as laid down in the statement of requirements.

In the context of a mechatronics environment such as that shown schematically in figure 1, once the allocation of system functions to the various domains has been achieved, by whatever means, the electronics designer is faced with a range of decisions which may or may not impact upon the decision made in relation to the other domains. A particular choice facing the electronics designer is that of the form of in which the required electronics functionality is to be achieved.

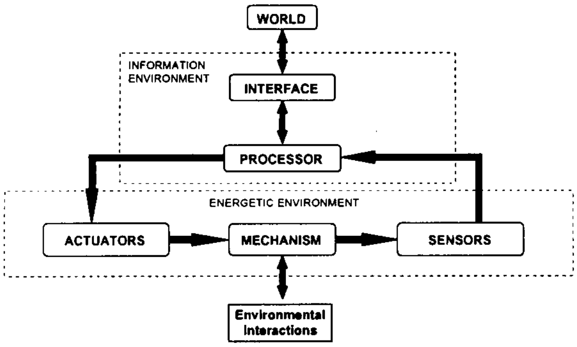


Figure 1. A mechatronics system

Among the primary choices facing the electronics designer is the form in which the circuit is to be implemented. For instance, for a particular application the choice might be between a microprocessor or microcontroller, a Digital Signal Processor (DSP) chip, an Application Specific Integrated Circuit (ASIC), a Field Programmable Gate Array (FPGA) or a Programmable Logic Device (PLD). Each of these is supported by its own set of design tools, often tailored to the particular device or family of devices in the case of FPGAs, PLDs and microprocessors/microcontrollers or towards a particular fabrication process in the case of ASICs.

Other areas that may influence choice include operating environment, production volumes, test and support facilities, ease of future modification and system development being neglected by concentration on the familiar and comfortable. Thus while production volumes may indicate a solution based on ASICs, the need for future upgrades may well suggest a software oriented solution based on microprocessors or DSP chips. It is also possible that a mixed signal analogue/digital solution is required rather than a purely analogue or digital system.

Faced with such a choice, particularly if under pressure to reduce the development time scale, the designer may well opt to choose that which is available and with which they are familiar as the basis for their design. While this strategy may well be successful in many instances it may well be mistaken in that a development of an existing solution may severely restrict the ability of the designer to act innovatively and to introduce new or novel solutions. In a mechatronic system it may in addition mean that opportunities are missed to further modify and simplify the mechanical elements of the design through an inappropriate choice of solution in the electronics domain.

3.1. Electronics design

In electronics design, designers have tended to follow a mixture of top-down, bottom-up and even middle-out strategies depending upon their familiarity with the problem and their past experience. In this way an experienced electronics designer may act to modularise the design while isolating those features of the system which are expected to be of significance in meeting the specified performance parameters. Once this modularization has been achieved then, as with software development, the individual modules can be built up and tested independently before connecting them to form a complete system.

While such an approach is often effective, particularly where the designer has long experience of the particular field, it may in some cases be time wasting or misleading, resulting in the generation of designs which are essentially minor modifications on previously existing designs.

What is required therefore is a mechanism which will support the designer in the partitioning of the design and the subsequent mapping of the resulting modules onto various forms of implementation and which will facilitate the choice of support and development tools by the designer.

3.2. Software design

In contrast to the above, the abstract nature of software has led to the development of tools such as Yourdon and HOOD as language independent processes and procedures to supporting the analysis and decomposition of the problem using what is essentially a top-down strategy.

The result is a modular approach to the design of software in which the process of decomposition is directly linked to the final code. Indeed, Computer Aided Software Engineering (CASE) tools often include a code generation feature in which the required code is automatically produced once the design is complete.

4. AN ELECTRONICS DESIGN ENVIRONMENT FOR MECHATRONICS

Within a mechatronics environment it is suggested that what is needed is a design environment that through its operation will support the partitioning of the design by enabling decisions as to the form of implementation to be used to be reflected in terms of its impact on other parts of the design and which will then facilitate the correct choice of support and development tools.

The Schemebuilder project in the Engineering Design Centre at Lancaster University is an example of such an approach applied to the design of mechatronic systems by providing support during the conceptual stages of the design process and enabling the designer to rapidly review alternative designs or schemes. However, within the context of any individual scheme the electronics designer is still faced with the problem of dealing with the diversity of solution forms available to them and of assessing the impact of a particular choice on the overall solution achieved.

Based on the previous comparison between the provision of support tools for software engineers and electronics design engineers it is argued that what is required is an equivalent of the CASE tools in the electronics domain. Such tools would support the high level decision making process associated with achieving a viable design and act to autonomously remove from consideration non-viable options.

In developing such a tool consideration must be given to the relative roles of the human designer and the tool. An evaluation of the capabilities of humans and computers will tend to suggest that these are to a significant degree complementary and that the computer should therefore be used predominately in those areas where humans are weakest such as large scale analysis. Thus while a human designer may well be capable of arrive at a viable solution as a result of intuitive reasoning they are less able to use techniques such as genetic algorithms which are much better suited to evaluation using a computer.

The concept of the proposed computer based tool the intent was therefore that the tool became a 'member' of the design team, making it essential that the strengths and weaknesses of the computer are integrated with the strengths and weakness of the human members of the team. The role of the computer is therefore to compliment the strengths and skills of the human designer.

4.1. The design environment

With the above concepts in mind it is now possible to put forward a structure for a computer based design support environment to support the design of the electronic elements of a mechatronic system. This proposed environment is shown in schematic form in figure 2 in which the central, shaded part represents the core design activities around which other activities are then grouped.

Referring to figure 2, the entry point to the environment if the Functional Decomposition module. This provide the designer with the ability to breakdown the system specification into its

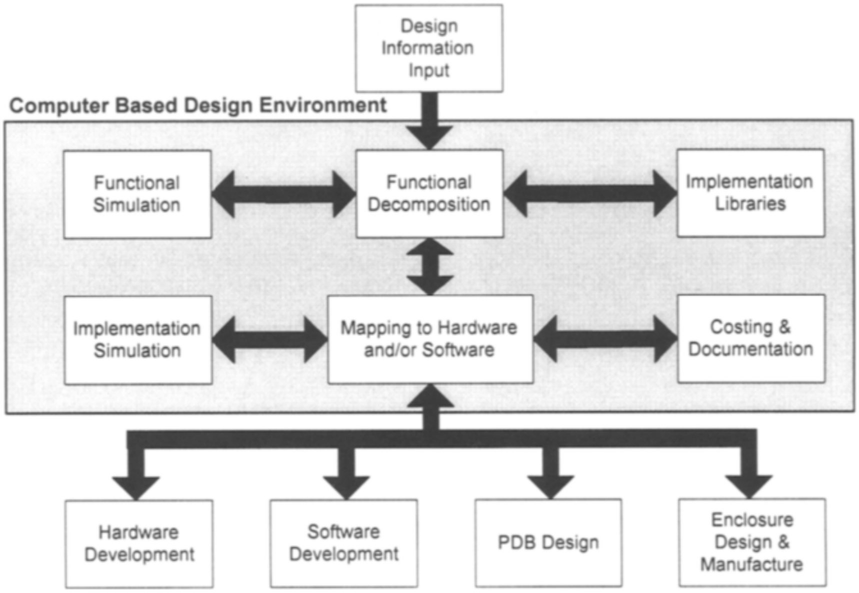

Design Information Input
Computer Based Design Environment
Functional Simulation
Functional Decomposition
Implementation Libraries
Implementation Simulation
Mapping to Hardware and/or Software
Costing & Documentation
Hardware Development
Software Development
PDB Design
Enclosure Design & Manufacture

Figure 1. Structuring the computer based design environment

functional elements and to evaluate the interaction of the individual functions through the use of functional simulation. This allows the designer to evaluate different solutions strategies and to identify key areas of the design such as time dependencies or noise sensitive areas.

As part of this process the functional decomposition is linked to the Implementation Libraries which contain prior solutions described in terms of their function and application. The Functional Decomposition module therefore attempts to map the current functional modules onto module held in the Implementation Library. Even if no full match is obtained, it is possible that a partial match can be found which can then form the basis of subsequent development either by modifications to other functions elsewhere in the system or by modification to the existing design. In the former case the chosen module would be inserted into the design and the system then used to identify the necessary changes elsewhere in the system while in the latter case the system would be used to identify the deficiencies in the current design.

Such an opportunistic approach is often used by human designers who extrapolate from existing solutions in the desired domain, or even from apparently unconnected domains. By using the computer to identify possible candidate solutions, the human designer is then left free to exercise their individual ingenuity.

Once a possible functional solution was obtained the system would attempt to map this to solutions in both the hardware and software domains while generating an operating simulation.

When a formal, or informal, mapping of the solution onto appropriate modules has been achieved then this can be evaluated using an Implementation Simulation module. This would typically be based around existing electronics simulators with some additional linking software to allow the results from one simulator be passed to another. Also at this stage information about costing would be introduced while a 'documentation generator' would be used to produce the necessary support material.

5. CONCLUSIONS

As the complexity of electronics and of mechatronic systems increases there is a need to provide support to the designer to enable them to rapidly evaluate different solutions and to assess their impact on the overall system. The proposed design environment is intended to support the electronics designer in carrying out these and other evaluations within the overall context of mechatronics following the initial decomposition and allocation of function to the various technologies.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Engineering and Physical Sciences Research Council in carrying out this work.

REFERENCES

1. Shah, V. & Gray, M., 1994, 'Structural processes for complex VLSI design', IEE Colloquium on Structured Methods for Hardware Design, p 8/1
2. Walters, R.M., 1996, 'The High-Level Design of Electronic Systems', PhD Thesis, Lancaster University

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A comparison between modelling approaches for custom electronic circuits within a control systems application

I.A. Grout[1], J. Santana-Corte[2], A. Winsby[1] and S.Burge[1]

[1] Engineering Department, Lancaster University, Lancaster, LA1 4YR, United Kingdom
[2] Electronics Department, ITESO University, Per. Sur 8585, Tlaquepaque, Jalisco 45090, Mexico

Within the design of mechatronic/control systems, there is an increasing use in system modelling and simulation techniques prior to the manufacture/fabrication stage. The choice on a suitable modelling approach can in many situations be based on experience in using a particular modelling/simulation tool. Where a system is to be derived by a designer who may have limited knowledge of the *pros and cons* of a particular modelling tool, the choice of an inappropriate tool may limit the effectiveness of this powerful design analysis aid. Additionally, where a design may be described at differing levels of abstraction, using a variation of possible terminology, the task of deriving a model from an abstracted description can be difficult. This paper will discuss a study into the modelling and simulation of a closed-loop control system using two modelling and simulation approaches. Benefits and limitations of both approaches will be discussed in reference to the target control system in order to identify the essential and generic features of these modelling approaches within the design of the custom electronics within a closed-loop control system.

1. Introduction

A number of approaches are possible in the design of the electronics to perform the sensor and actuator interfacing in addition to the algorithm generation within typical closed-loop control systems[1][2], see figure 1. These are :

1. Analogue electronics (hardware)
2. Mixed-signal electronics, microprocessor/microcontroller (hardware-software) algorithm implementation
3. Mixed-signal electronics, custom integrated circuit solution (ASIC) algorithm implementation (digital core)

Method 2 has gained a great deal of success in the ability to define complex algorithm structures in software whilst maintaining the ability for ease of reconfiguration. In addition to the software implementation of digital control algorithms, custom Integrated Circuits (ASICs : Application Specific Integrated Circuits)[3][4] may provide an increasingly attractive solution to the electronics implementation by providing the ability to implement complex mixed-signal functions within a compact package. Coupled with the increasing complexity of the overall control system, the need to consider and implement modelling approaches to

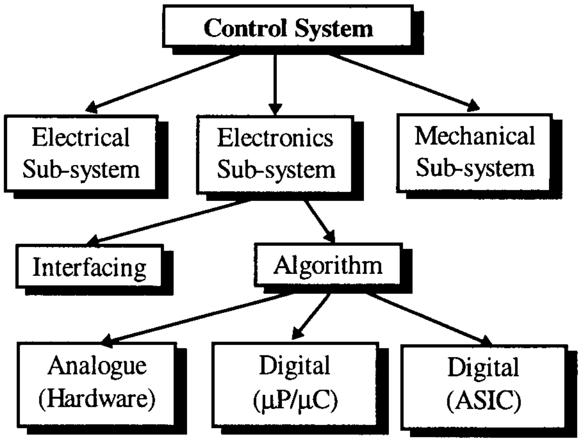


Figure 1 : Solutions to Algorithm Implementation

predict the operation of the system prior to manufacture/fabrication is required. This is true for the electronic sub-systems as it is for the mechanical and electrical (power) sub-systems. A number of tools are available to aid the designer in the modelling and simulation task. These Computer Aided Control System Design (CACSD)[5] packages such as MATLAB/SIMULINK[6][7] can provide useful analysis of the system design but may lack in detailed consideration of the need to extract information from the models in order to generate the electronics hardware. This is particular relevance to mixed-signal ASIC designs where the designer has a great deal of flexibility in the architectures that can be generated within a small physical size product, which can trade-off hardware against software based actions and serial operation of events with parallelism. Approaching the problem from the electronics design/analysis perspective, a modelling and simulation approach needs to consider the modelling techniques from the control perspective as well as the underlying circuit electrical behaviour. Hardware Description Languages (HDLs) allow the user to generate text based models for (primarily) the electronics starting at a high level (*behavioural*) description before considering detailed (*structural*) descriptions which can map to the intended circuit. Both CACSD and HDLs are considered here in relation to closed-loop control systems. A number of benefits and limitations of both approaches are discussed in reference to the target control system in order to identify the essential and generic features of these modelling approaches within the design of control systems.

The starting point for this study is the design specification which, for this study, is in the form of a requirements document. The case study is based around a single-input, single-output (SISO) closed-loop motor position control system. The study targets the electronics sub-system within the overall control system and it is assumed that the mechanical sub-system has been suitably modelled and is a predetermined input. Similarly, the electrical sub-system (electric motor, power drive circuitry) is also a predetermined input. The requirements document will specify the options available to the designer in terms of hardware availability in order to derive the electronics sub-system functionality.

2. Modelling Approaches

The approach chosen was to develop and simulate MATLAB/SIMULINK[6][7] and A-HDL[8][9][10] models for the control system. Starting from the high level description of the overall design specification, the electronic sub-system under consideration was developed, modelled and simulated, see figure 2. Although the initial work considered the three potential implementation methods for the electronics, the study targeted the custom designed, controller ASIC based method. From this, the type of information and presentation format are highlighted since, in general, the requirements document will not necessarily contain all relevant information on the detailed implementation issues for the device. The development of the models will determine the usefulness of this initial document.

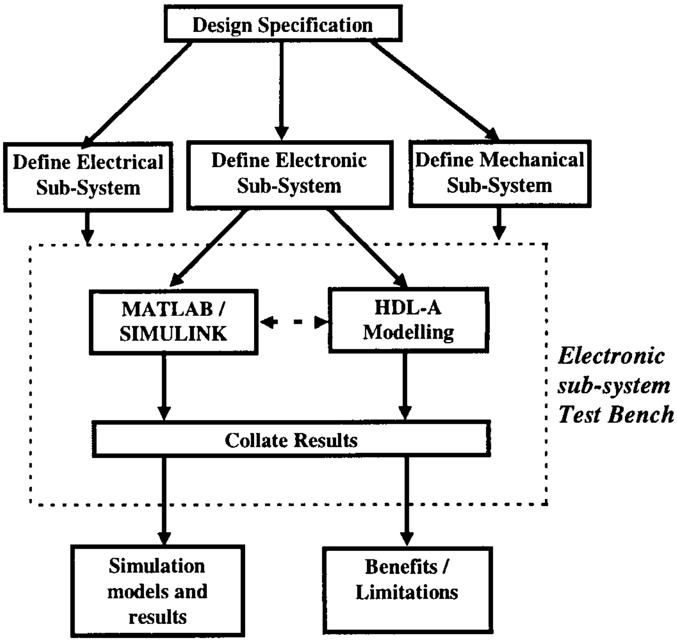

Design Specification
Define Electrical Sub-System
Define Electronic Sub-System
Define Mechanical Sub-System
MATLAB / SIMULINK
HDL-A Modelling
Collate Results
Electronic sub-system Test Bench
Simulation models and results
Benefits / Limitations

Figure 2 : Simplified study flowchart

The flowchart in figure 2 identifies a simplified flow for the study. MATLAB/SIMULINK models have been compared to an A-HDL model by exercising the design through a set of suitable test stimuli. The control system has been partitioned into the electronic, electrical and mechanical sub-systems. All have a relevant part within the *electronic sub-system test bench* which will define the test vector application and results collation. To provide for a study that has a common basis on which to base conclusions, the following list provides a (by no means comprehensive!) starting point for initial discussions on issues within the overall study flow,

- Electronic sub-system partitioning
- Optimisation of models
- Abstraction level
- Mixing of functional and structural descriptions
- Verification of models
- Setting-up the test bench
- Stimulus application
- Collation of modelling and simulation results
- Presentation of results

In many situations, the choice of model abstraction level will determine the usefulness and applicability of the results. Simple models may provide speed advantages but with a loss of accuracy under certain operating conditions, for example, with the analogue electronics driven into saturation or the response to high frequency input signals. The study identifies a number of these aspects in relation to the model definition of the electronics sub-system.

3. Digital Electronic System Modelling using HDLs

The modelling of digital electronic circuits and systems using an appropriate hardware description language (HDL) is common-place in the design of digital microelectronic products. Here, the HDL is used to describe the design at a number of possible levels of abstraction. High level *behavioural modelling* allows the designer to develop and debug the behavioural operation of the circuit/system before a fabrication route is chosen, see figure 3. This initially allows the design to be technology independent and allows the designer to concentrate on the functionality issues before the specific implementation issues. Additionally, at the early stages of a complex system design, the implementation route may not yet be known. In this case, the behavioural modelling approach would be the only way forward. The two standard HDLs currently in use are VERILOG[11] and VHDL[12]. Both these are used in the modelling, simulation and synthesis of designs.

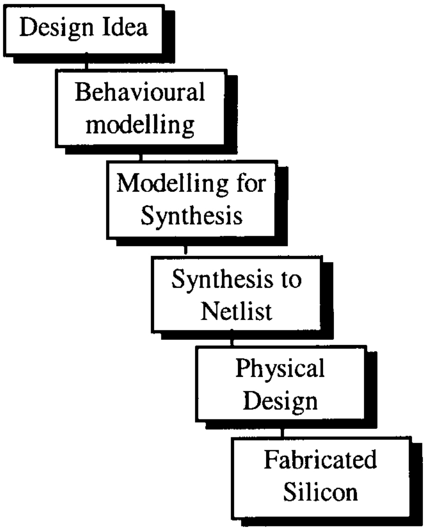


Figure 3 : Digital HDL Synthesis

Whilst digital HDLs allow for the digital part of the overall system to be developed (and ultimately synthesised into a circuit netlist capable of being fabricated), the remainder of the system is missing from this design stage. Moves towards analogue HDLs (AHDLs) are now allowing for the remainder of the system to be modelled and simulated along with the digital portion. These also allow for the electronic and non-electronic system parts to be modelled and simulated in a single language.

4. Control System Study Design

The control system is based around the position control of a DC electric motor module. The motor is a 12V DC motor and is housed in a module containing electronic voltage and power amplifiers. The signal input is an analogue voltage in the range -5V to +5V, see figure 4. This is amplified internally to the module to -12V to +12V. The module output is a position which ranges 0^0 to 359^0. The position is sensed by a potentiometer which provides an output in the range -5V to +5V (0V = 180^0). The motor speed is sensed by a

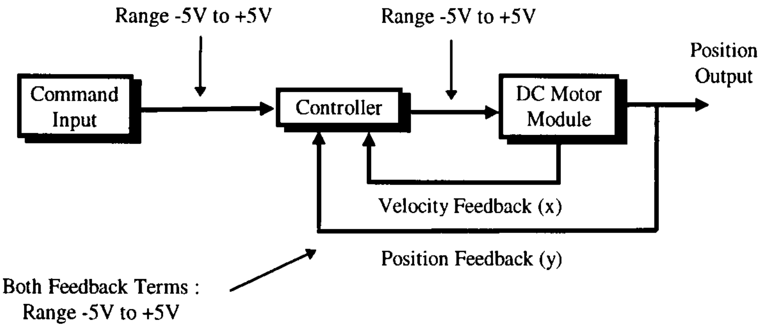


Figure 4 : Motor Position Control

tachogenerator which provides an analogue output voltage in the range -5V to +5V (0V = motor stationary). The Controller ASIC used operated on a +5V power supply and required signal conditioning circuitry to match this to the motor module voltage range, i.e. -5V to +5V.

5. SIMULINK Modelling

The block-diagram approach to model generation adopted within SIMULINK lends itself well to a top-down approach linking the different technologies within the control system. Here, MATLAB *version 4* and Simulink *version 1* were used. However, the modelling approach required simplification of the detailed operation of the electronics (for example, digital action timing and quantisation effects) which in this case could be accepted. Figure 5 shows the top level model identifying the Controller ASIC and DC Motor Module plant along with the required signal conditioning circuitry, matching the ASIC voltage levels to the Motor Module. These input (output) op-amp based circuits consist attenuation (amplification), 2^{nd} order low-pass anti-alias filtering and level shifting. The ASIC was modelled by its' algorithm structure with quantisation and timing effects grouped together within the input and output sub-blocks interfacing the control algorithm to the top level control system.

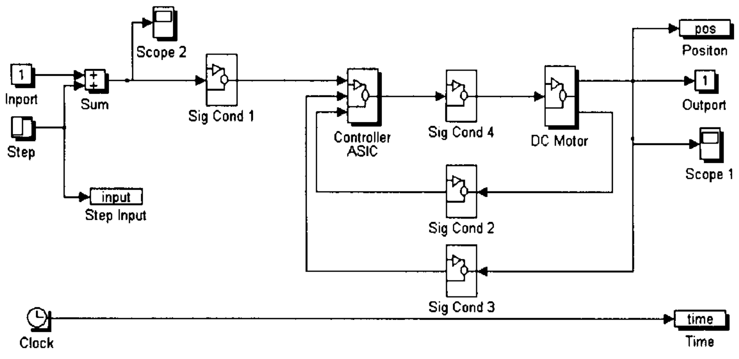


Figure 5 : Motor Position Control, SIMULINK Model

6. HDL-A Modelling

HDL-A is a language used to develop models of analogue and mixed-signal circuits. The same language syntax is used to develop models for the non-electronic parts of a mechatronic system. HDL-A is based on VHDL syntax and semantics and utilises extensions to model the analogue parts and a sub-set of VHDL to model the digital parts. The models are essentially analogue behavioural models and allow for both time domain (transient) and frequency domain (ac) analysis. One major advantage of using this modelling approach is (in digital HDL terminology) the ability to mix *behavioural* and *structural* levels of description. This allows the user to initially develop simplified behavioural models to test ideas before expending the effort to develop detailed models detailing the circuit structure. Ideally, at some point in the process, it would be ideal to be able to synthesise this *text description* into actual hardware. This is current practice for digital designs (and in some cases, for software code generation) but for analogue and mixed-signal hardware designs, particularly mixed-signal ASICs, a formal, structured approach has yet to be established. Current work on HDL-A is a step in the right direction to future advances. Figure 6 shows the generic structure of the HDL-A code. The HDL-A code is developed and compiled before simulation. In this approach, Accusum II©[13] is used to simulate the compiled models. Following the concepts of

VHDL, a design is created with an *Entity* and *Architecture*. The *Entity* of a design provides an interface to the external world that it is connected to, so defining connectivity and any generic declarations. The *Architecture* of the design defines the detailed operation of the design for initial, ac, dc and transient conditions. Analysis commands entered into AccusimII are then used to identify the system response.

```
ENTITY hdla_model IS
        GENERIC (params_list : data type):
        PORT (dig_terminal_list : data_type);
        COUPLING (states_list : data_type; . . .);
        PIN(pin_list : nature; . . .);
END ENTITY;


ARCHITECTURE arch_name OF hdla_model IS
        SUBPROGRAM [declare C program];
        CONSTANT [constant_names : data_type);
        VARIABLE [variable_list : data_type];
        SIGNAL [signal_list : data_type];
        STATE [state_list : data_type];
        BEGIN
                RELATION
                PROCEDURAL FOR INIT =>
                        [initialize parameters in GENERIC statement];
                        [set up initial condition values];
                PROCEDURAL FOR ac, dc, transient =>
                        [behavior description can break up into individual procedurals];
                EQUATION (unknown) FOR analysis_type_names =>
                        [equation using the "==" asignement];
        END RELATION;
        PROCESS
                BEGIN
                        [digital initializing statements];
                        [optional_label] : LOOP
                WAIT    statement; --only one WAIT in PROCESS
                        [sequence of digital statements];
                END LOOP;
        END PROCESS;
END ARCHITECTURE arch_name;
```

Figure 6 : Generic HDL-A Model Structure

7. Model Simulation

7.1 Introduction

The simulation study is based around the ability to generate a suitable model and to run both a transient (step input) and frequency response (gain and phase) for the controller with the following coefficients as defined in table 1, where b, kp, ki, kd, kv and kfp are action coefficients and T the sampling period for the particular ASIC design used in the study. Other controller and control system strategies would have different requirements and so the two approaches adopted would need further consideration. Additionally, synthesis potential was considered as an extension to the simulation work undertaken.

7.2 Transient (Step) Response

With both approaches, the transient analysis was straight forward to run and produced no problems in obtaining useful results. With SIMULINK, the results were plotted from the MATLAB workspace with the time and input(*input*)/output(*pos*) variables. With the HDL-A results, AccsimII commands were used after the analysis of the transient behavioural models. As an example, figure 7 shows the step response from the SIMULINK

Coefficient	Value
b	1
kp	5
ki	5
kd	0
kv	3.75
kfp	0
T	0.01 seconds (digital controller only)

Table 1 : Controller coefficient values

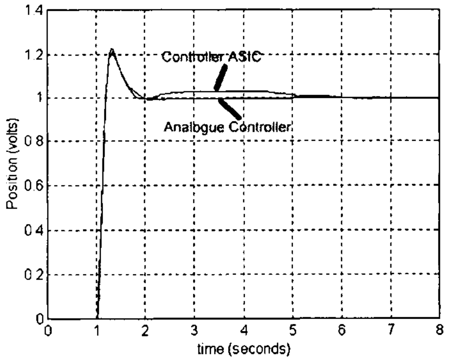


Figure 7 : SIMULINK Model, Step Response

model which is compared here to the analogue equivalent controller.

7.3 Frequency Response

The frequency response initially caused a number of problems. The need to extract a state-space model into MATLAB from the SIMULINK model required simplification of the models to enable a linear state-space model to be extracted. This was a limitation with the approach taken and further distanced the models from the implementation means. It also required an understanding into how the state-space model was created and how to manipulate the design to enable results to be obtained. With the HDL-A model, this was less problematic as the frequency response was defined in addition to the transient response

7.4 Synthesis capabilities

From the electronic circuit design perspective, and taking into account the effort expended by creating and verifying the models, it is useful to have a further use for these models created. Synthesis of the complete electronic circuit model would be a major step forward, although this is not as yet possible. However, with the concept of HDL-A coming from the use of VHDL, it may be considered that future developments would be in the ability to synthesise the model into hardware (and where appropriate, software), for the digital, analogue and mixed-signal functions. This may have more chance of success with its background in the digital VHDL domain than a control oriented CACSD package, although this is far from certain. Ease of use and robustness of model translation would be required and currently, with the block diagram approach of SIMULINK, this would be a more intuitive approach from the control system design perspective.

8. Conclusions

This paper has discussed the use of two modelling approaches for the electronics sub-system design in a closed-loop control system. Traditional CACSD using MATLAB/SIMULINK was compared to the HDL-A modelling language. HDL-A is a development from the industry standard digital VHDL which are used to model, simulate and synthesise descriptions of the digital circuits/systems currently developed within the microelectronics industry. Where the aim is to implement a design within electronics hardware, the range of analysis and simulation tools available will be based on the end user background. For example, MATLAB/SIMULINK comes from a control systems background whereas HDL-A from an electronics/microelectronics background. Each approach provides benefits and limitations.

9. References

1. Astom K. and Wittenmark B., *Computer-controlled systems theory and practice*, Second edition, Prentice Hall, ISBN 0-13-172784-2, 1990
2. Jacob, J., *Industrial Control Electronics, Applications and Design*, Prentice-Hall International, USA, 1989
3. Grout I, Burge S and Winsby A., *Mixed-signal custom IC control processors incorporating design for test/self-test*, IEE Colloquium on DSP Chips, 25th September 1997, pp 3/1 - 3/8
4. Winsby, A. Burge S. and Grout, I., *The Design and Implementation of a Surface Oriented Controller*, Mechatronics 98, September 1998
5. Jamshidi M. et al, *Computer-aided analysis and design of linear control systems*, Prentice Hall, ISBN 0-13-151796-1, 1992
6. *MATLAB*, The Math Works Inc., USA
7. *SIMULINK*, The Math Works Inc., USA
8. Binns R. et al, *High-level design of analogue circuitry using an analogue hardware description language*, IEE Colloquium on Mixed-Signal AHDL/VHDL modelling and synthesis, !9th November 1997, pp 3/1 - 3/8
9. Wilton R., *Developments in and applications of mixed-signal HDL tools*, IEE Colloquium on Mixed-Signal AHDL/VHDL modelling and synthesis, !9th November 1997, pp 3/1 - 3/8
10. HDL-A, ANACAD Computer Systems, GmbH
11. Blair, G., *Verilog: accelerating digital design*, IEE Electronics and Communication Engineering Journal, April 1997, pp68-72
12. Navabi Z., *VHDL Analysis and Modeling of Digital Systems*, Mc-Graw Hill International Editions, 1993, ISBN 0-07-112732-1
13. Accsuim II, Mentor Graphics Corporation, USA

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Wireless Remote Sensing and Identification by Using Passive SAW Devices

Werner Buff and Michael Goroll

Technical University of Ilmenau, P.O. Box 100565, D-98684 Ilmenau, Germany

Abstract

Surface acoustic waves (SAW) have been employed in a large number of different devices mainly for frequency control applications. Today the annual production of SAW devices world-wide is more than 200 million pieces. Every TV-set manufactured today uses at least one SAW filter in the IF stage of the receiver. Moreover, there are a lot of other applications as well, and the number of devices produced is increasing rapidly. The main reasons for that are their small sizes, low costs and high performance due to utilising the technological improvements of the well-developed microelectronic technique. Furthermore, they are highly reliable and reproducible due to their solid-state properties and the relatively good understanding of the basic physical processes.

From a physical point of view, SAW are mechanical waves travelling along the surface of a solid. So called interdigital transducers (IDT) overlaying as a specially shaped metallic structure on the surface of a piezoelectric crystal usually are employed for the excitation. An IDT is made by a photolithographic process and is the key element for all SAW devices. By using the inverse piezoelectric effect such a structure excites SAW with a velocity of about 10^5 times smaller than the velocity of electromagnetic waves in vacuum. For this reason it is possible to produce SAW devices in a frequency range from 10 MHz up to some GHz. Depending on the crystal cut and the direction of the travelling wave different types of waves do exist. Thus SAW can be distinguished in different types depending on the particle displacement (linear, elliptical or circular polarisation), on the attenuation function from the surface into the volume of the substrate, on the coupling to an electric field or not and the material system where waves are travelling (monocrystaline, crystal with or without piezoelectric behaviour, layered structures, membrane). For frequency controlling devices there are almost always used Rayleigh-type waves with an elliptical particle displacement. The most frequently used materials are Quartz (SiO_2) or Lithiumniobate ($LiNbO_3$), respectively, in monocrystaline form.

The mainly used type of frequency filters is a delay line with two IDT's and a travelling path for the SAW between. The behaviour of such a construction is determined by material parameters, the shape of the two IDT's an the physical conditions for the travelling SAW. There is a great variety of different filter types meeting the demands of the electronic industry. The limits of today's filters are approximately in a range from 10 MHz up to 2 GHz with a fractional bandwidth up to 20% and bandpass characteristics corresponding to the customers specifications.

The second main type of SAW devices is the resonator with one or two IDT's and reflecting gratings next to them. The case of resonance means that only standing waves are existing and the sharp resonance peak in the frequency is only depending on the geometrical layout of the exciting IDT and the gratings in connection with the propagation conditions for the SAW.

Devices used for signal processing must have a high stability and have to be insensitive to variations of environmental conditions. However, this is not possible. It is known that SAW devices react to changes in physical conditions in the propagation path by changes both their amplitude and phase velocity. A variation of the phase velocity leads to a frequency shift and the change in amplitude can be measured immediately. So it is possible to make use of such changes in the electric characteristics for „feeling“ changes in physical environmental conditions.

SAW sensors have the advantages of extremely high sensitivity and an output signal in the form of a frequency shift which enables them to be directly connected to a digital signal processing without any analogue/digital conversion. Moreover, there is a possibility to construct passive sensors where the sensor element itself does not need any energy supply. The principle is based on the big dynamic range for SAW excitation and detection. If one of the IDT's of a SAW device is connected to an antenna then it possible to excite SAW via a RF signal transmitted from an interrogating equipment and received by the antenna. The SAW excited in this way travels along the surface and is influenced by environmental conditions. Thus the response of the device retransmitted to the interrogation equipment by the same antenna contains an information about the environment (e.g. temperature, pressure, etc.). As an example, the SAW structure can be a resonator. An interrogating RF signal can create a standing wave in this resonator element. The resonance frequency in the moment of measurement is influenced by the environmental conditions. After the interrogating signal is switched off the resonator oscillates with attenuated oscillations with the actual resonance frequency. This attenuated oscillation is retransmitted via the antenna to the interrogation equipment and can be processed. Another possibilities are to use reflectors on the surface of the element or to design the delay line in the form of a tapped delay line.

In a similar manner it is possible to construct passive identification tags. They also do not need any own energy source and can be interrogated in a wireless manner via a RF signal as well.

When the principles are combined in integrated or hybrid form there are a lot of new possibilities to get information from places which are not or very hard accessible otherwise. Hereby the sensor together with the amount of a physical parameter simultaneously gives an identification signal and so the place can be identified where the measurement occurs.

A whole string of examples shows the capability of this concept.

1. INTRODUCTION

In the recent years surface acoustic wave resonators (SAWRs) have proved to be excellent high Q sensing elements. The advantages of SAWRs such as high sensitivity, accuracy, long-term stability and the possibility of storing electromagnetic energy, allow them to be used in passive remote sensor systems [1-3]. Fig. 1 shows the fundamental structure of an one-port SAWR and the behaviour of an example with a centre frequency of f_0 = 433.347 MHz in frequency domain. Due to the inverse piezoelectric effect the interdigital transducer (IDT) is generating surface acoustic waves (SAWs) which are reflected by the refector arrays. After the reflection the SAWs are detected by the IDT (piezoelectric effect) thus creating standing waves.

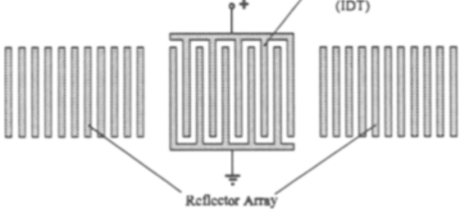


Figure 1a. One-port SAWR -structure

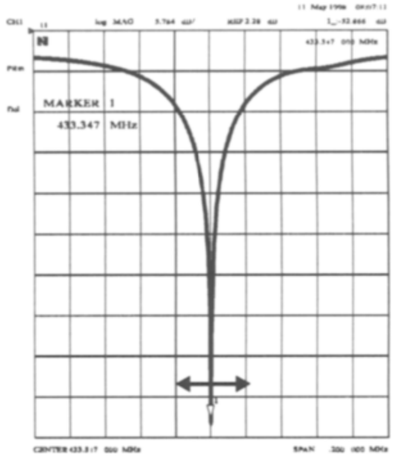


Figure 1b. One-port SAWR - structure S_{11}-measurement for an example (f_0 = 433.347 MHz)

2. DIFFERENCE MEASUREMENT METHOD

SAWR sensors are often employed as sensor devices in passive wireless telemetry systems because of the high centre frequency, low cost and easy connection to an antenna. In [3] a passive wireless remote sensor system is described.

Measurements with single SAWRs in telemetry systems showed a high radio channel influence on the resonance frequency in this network. To minimize this influence the difference measurement method with two SAWRs has been choosen (see Fig. 2) [3].

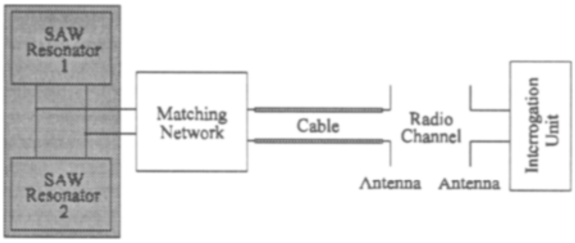


Figure 2. Passive wireless SAWR telemetry system [3]

For such a system (f_0 = 313.7 and 313.9 MHz, antenna distance: (0.2...0.7) m) a maximum shift of the difference frequency f_D of only 0.401 kHz was measured (see Fig. 3) [3].

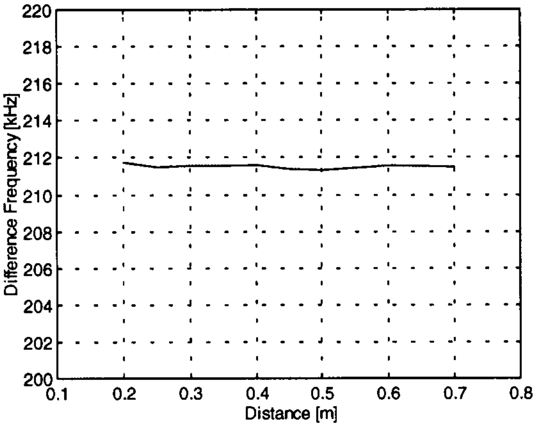


Figure 3. Difference frequency of a two-SAWR sensor system in dependence of the antenna distance [3]

Due to the differential measurement method the exact determination of the property change which is to be detected with almost no influence of the radio channel is possible. Furthermore, the using of two sensing elements in the same sensor leeds to possibilities to influence the final sensor sensitivity. For other than temperature sensors it is possible to realize a good sensor temperature compensation.

3. SAWR SENSORS FOR PASSIVE REMOTE SENSING

The basic arrangement for the described SAWR sensors is the structure in Fig. 4. It consists of a quadratic plate of Y-35.5°-cut-Quartz (18×18 mm). The plate was produced by *TfT Telefilter Teltow*, Germany. The SAWRs are metallic configurations of Al-metallization (150 nm) structured by a microelectronic lithographic process and cutted by the microelectronical saw-technique.

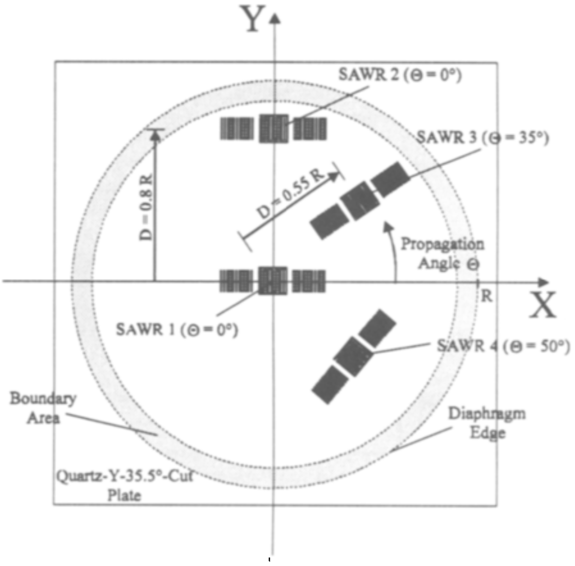


Figure 4. Teststructure of SAWR sensors on Y-35.5°-cut-Quartz

Using the anisotropic behaviour, the SAWRs are characterized by different locations on the plate and different SAW propagation directions (propagation angle Θ). A circle (centered on the plate) with a radius R is the basis of the SAWR-arrangement.
The centre frequency is influenced by temperature T and pressure p (see Eq. 1, 1a, 1b) [4].

$$f(p,T) \cong f_0 + S_p^*(p-p_0) + TC^*(T-T_0)] \qquad (1)$$

$$S_p^* = f_0 S_p \quad \text{and} \qquad TC^* = f_0 TC \qquad (1a), (1b)$$

T_0 and p_0 are reference values, S_p is the pressure sensitivity and TC the temperature coefficient of frequency [4,5].

3.1 TEMPERATURE SENSOR

The dependence of the *TC* from the propagation angle Θ is the basis for the design. Fig. 5 shows the theoretical and measured curves for the Y-35.5°-cut-Quartz [4,6]. They differ by 20 ppm/K. The inaccuracy of the production of the quartz plate and the fotolithographic mask as well as the *TC*s of the matching network elements are the reasons for this difference [4].

To produce a temperature sensor SAWRs with different *TC*s are necessary.

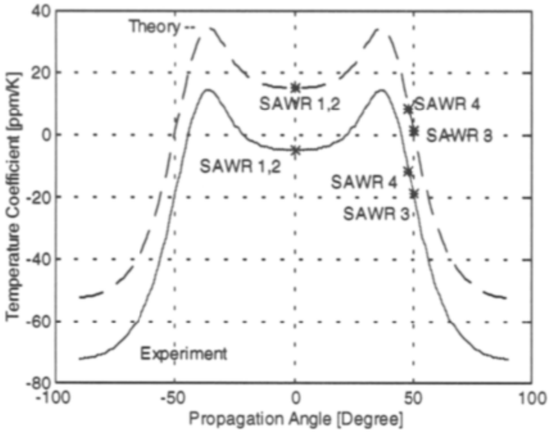


Figure 5. Temperature coefficient of frequency versus the propagation angle for Y-35.5°-cut-Quartz

SAWR	TC [ppm/K] Theory [6]	TC [ppm/K] Experiment
1	+15	-1
2	+15	-1
3	+34	+18
4	+1	-16

Table 1. *TC*s of the teststructure SAWRs

Tab. 1 presents the *TC*s of all SAWRs. SAWR 1 and 3 are choosen to form the temperature sensor. Fig. 6 shows f_D of SAWR 3 and 1 in a range of (30...75) °C. The measured difference frequency f_D has a *TC* of +19 ppm/K (+8.3 kHz/K).

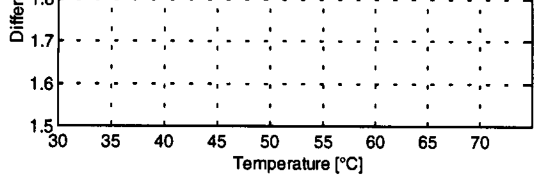


Figure 6. Difference frequency of SAWR 3 and 1 ((30...75) °C)

3.2 PRESSURE SENSOR

To build a pressure sensor two SAWR with different pressure sensitivities S_p are necessary. In the same way SAWRs with identical *TC*s must realize a good temperature compensation.

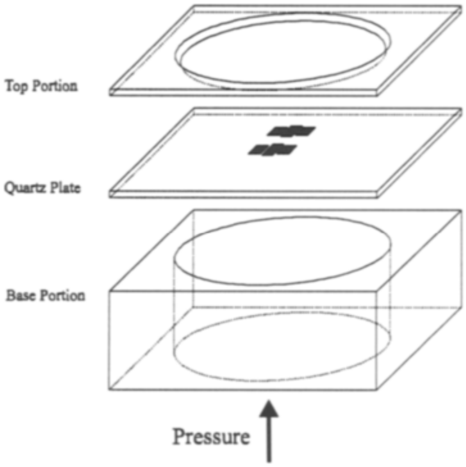


Figure 7. Structure of the realized pressure sensor

Fig. 7 shows the structure of the pressure sensor. The Quartz plate is circular clamped between a base and a top portion (both steel) and sealed by a rubber ring. The pressure acts from the back site on the SAWRs. The pressure sensitivity of any point on the diaphragm (characterized by the distance *D*) was calculated by Eq. 2 [4, 6-10].

$$S(D) = k_0 \frac{3(1-\mu_S^2)}{8E} \left(\frac{R}{h}\right)^2 \left(k_1 + k_2 \left(\frac{D}{R}\right)^2\right) \qquad (2)$$

The diaphragm parameter are shown in Tab. 2.

Parameter	Value	
Poisson Ratio μ_S	0.14	[11]
Young´s Modulus *E*	76 GPa	[12]
Radius *R*	6500 microns	-
Thickness *h*	500 microns	-

Table 2. Used Quartz diaphragm parameter

SAWR	k_0	k_1	k_2
1, 2	1	-1.085863	+1.494
3, 4	0.9218	-0.791412	+3.396

Table 3. Constants k_0, k_1 and k_2

The pressure sensitivity S_p has been calculated by using the constants k_0, k_1 and k_2 which are presented in Tab. 3. The constants depend on the material, the propagation direction and the geometry of the SAWRs [4, 6-10]. The analysis shows that the constants k_0, k_1 and k_2 in our case for SAWR 3 and 4 are approximately identical. The S_p-values of all SAWRs in dependence on the position D of the diaphragm are shown in Fig. 8.

The measurements show that a statement of a concrete package is very important. An boundary area is necessary because the geometry of every SAWR must take into consideration near the diaphragm edge.

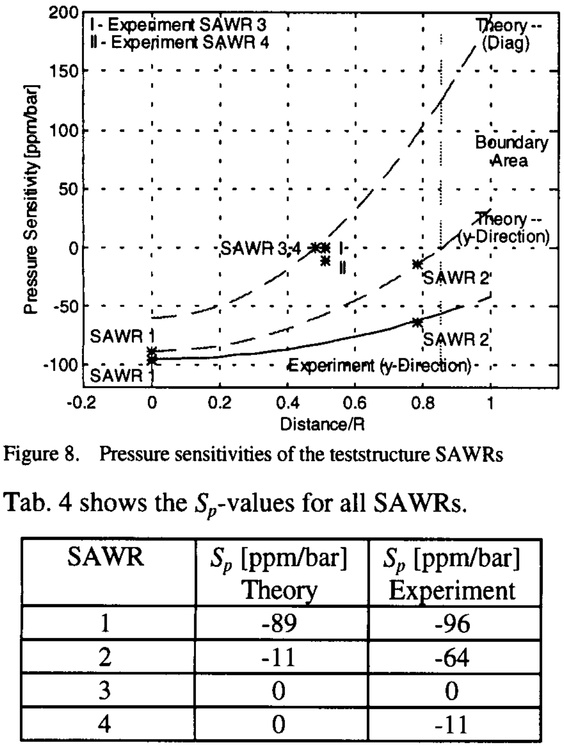


Figure 8. Pressure sensitivities of the teststructure SAWRs

Tab. 4 shows the S_p-values for all SAWRs.

SAWR	S_p [ppm/bar] Theory	S_p [ppm/bar] Experiment
1	-89	-96
2	-11	-64
3	0	0
4	0	-11

Table 4. Measured and calculated S_p-values

To realize the pressure sensor SAWR 1 and 2 of Fig. 4 were choosen. In Fig. 9 the pressure behaviour of f_D is plotted (pressure range: (0...6) bar). An experimental pressure sensitivity of -32 ppm/bar (-14 kHz/bar) was reached.

There is some deviation from the theory which is connected with a not sufficient stable experimental equipment. But the diagrams show also a principal agreement.

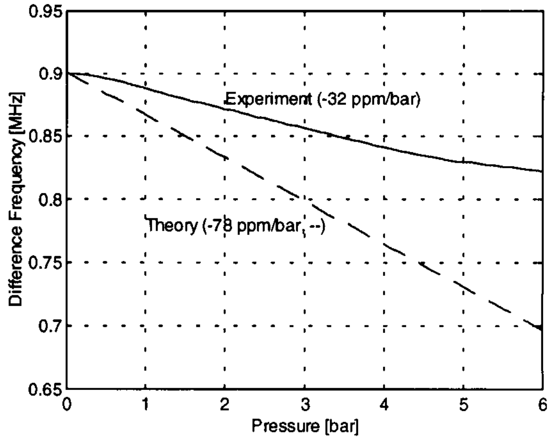


Figure 9. Difference frequency for a pressure range of (0...6) bar

3.3 PRESSURE-TEMPERATURE SENSOR

Using the results of chapter 3.1 and 3.2 a combined pressure and temperature sensor was realized. According to Fig. 2 a sensor system with three SAWRs was constructed. Now in case of remote measurement, the interrogation unit has to analyze two difference frequencies. After a calibration the interrogation unit calculates the predominant temperature and pressure according to Eq. 3 [4].

$$p = f(f_{D21}, f_{D32}), \qquad T = f(f_{D21}, f_{D32}) \qquad (3)$$

Using the results of Tab. 1 and 4 the SAWRs 1, 3 and 4 were choosen. Fig. 10 and 11 show the pressure and temperature behaviour of the difference frequencies (T = (0...2) °C, p = (0...2) bar).

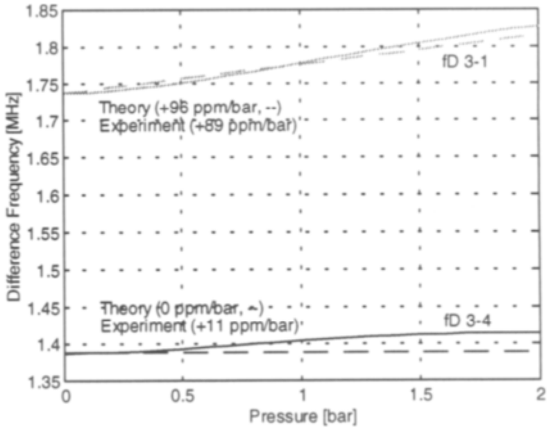


Figure 10. Pressure behaviour of the difference frequencies ((0...2) bar, (30...75) °C)

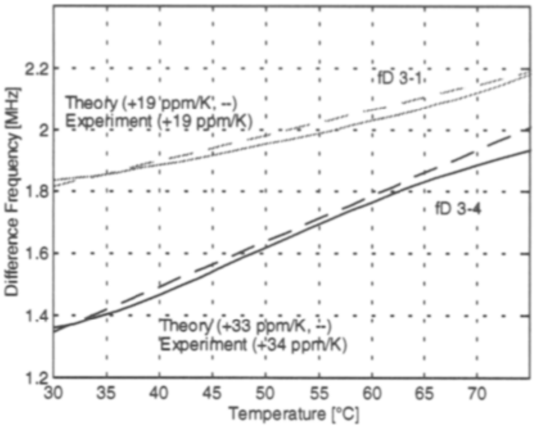


Figure 11. Temperature behaviour of the difference frequencies ((0...2) bar, (30...75) °C)

The experimental results for f_{D31} were +89 ppm/bar (+39 kHz/bar) and +19 ppm/K (+8 kHz/K), f_{D34} has measured values of +11 ppm/bar (+15 kHz/bar) and +34 ppm/K (+15 kHz/K).

4. IDENTIFICATION

Just as remote sensor systems the application of identification devices (ID-tags) in remote sensor systems requires the use of released frequency ranges. As an example, the European limitations in the ISM-band (433.92 ± 0.87 MHz) are too strict to be covered by conventional SAW-ID-tags based on reflection structures or tapped delay lines [13]. Both types of devices suffer from the broad bandwith of the short impulse that are used to excite acoustic waves on them. To overcome this problem, a special SAW device design, derived from conventional tapped delay lines and optimized for wireless interrogation systems as a two-terminal ID-tag was developed (see Fig. 12) [14]. This new design allows low-bandwidth and low-power interrogation using long, unmodulated sine bursts. The device responds with a PSK-code sequence suitable for a new wireless CDMA-based identification system that allows a simultaneous interrogation of multiple ID-tags by using orthogonal codes.

The working principle is based on superposition of the correlation signals of multiple apodized transducer pairs that are connected in parallel. In the device, every transducer (tap) works as both, receiver and transmitter of surface acoustic waves.

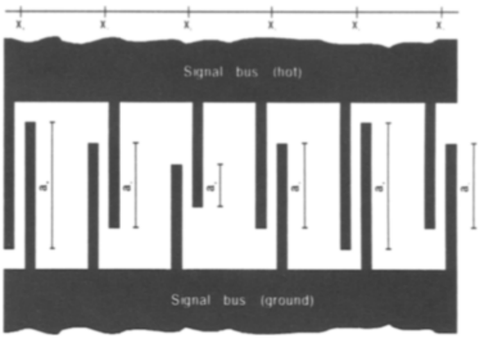


Figure 12. Identification device after [14]

REFERENCES

1. Buff, W., "SAW Sensors", Sensors and Actuators, (17) 1989, pp. 55-66
2. Buff, W et. al.: "Remote Sensor System using Passive SAW Sensors", *Proc. IEEE Ultrasonics Symp.*, 1994, pp. 585-588
3. Buff, W. et. al., "A Difference Measurement SAW Device for Passive Remote Sensing", *Proc. IEEE Ultrasonics Symp.*, 1996, pp. 343-346
4. Buff, W. et. al., "Universal Pressure and Temperature SAW Sensor For Wireless Applications", *Proc. IEEE Ultrasonics Symp.*, 1997, pp. 359-362
5. T.M. Reeder, D.E. Cullen and M. Gilden, "SAW Oscillator Pressure Sensor", *Proc. IEEE Ultrasonics Symp.*, 1975, pp. 264-268
6. Weihnacht, M., "Berechnungen für BMBF-Projekt SENSIDENS", Dresden, 1997
7. Cullen, D. E. et. al., "Progress in the Development of SAW Resonator Pressure Transducers", *Proc. IEEE Ultrasonics Symp.*, 1980, pp. 696-701
8. Cullen, D. E. et. al., "Measurement of SAW Velocity Versus Strain For YX and ST Quartz", *Proc. IEEE Ultrasonics Symp.*, 1975, pp. 519-522
9. Taziev, R. M. et. al., "Pressure-Sensitive Cuts for Surface Acoustic Waves in α-Quartz", *IEEE Trans. on Ultrasonics, Ferroelectrics and Freq. Control*, 1995, pp. 845-849
10. Goroll, M., "Drahtlos fernabfragbare Drucksensoren auf der Basis von SAW-Bauelementen", Technical University of Ilmenau, *Diploma Thesis*, Ilmenau, 1996
11. Neumeister, J., "Druckmessung mit Oberflächenwellenfiltern sowie resistiven und kapazitiven Dünn- und Dickschichtsensoren", University of Stuttgart, *Ph.D*, Stuttgart, 1991
12. Wagner H.-J., "Entwicklung von Technologien zur Herstellung piezoelektrisch angeregten, mikromechanischen Resonatorstrukturen in Silizium und Quarz", Technical University of Braunschweig, *Ph.D.*, Braunschweig, 1995
13. Reindl, L. et. al.: "Programmable Reflectors for SAW ID-Tags", *Proc. IEEE Ultrasonics Symp.*, 1993, pp. 125-130
14. Vandahl, T.: "Mehrfach gewichtete, selbstkorrelierende Verzögerungsleitung auf der Basis akustischer Oberflächenwellen für passive Identifikations- und Sensoranwendungen", Technical University of Ilmenau, *Ph.D.*, Ilmenau, 1997

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Wireless Identification and Sensing Using Surface Acoustic Wave Device

Dr. A. Springer[a], O.-Univ. Prof. Dr. Ing.habil. R. Weigel[a], Dr. A. Pohl[b] and Univ. Prof. Dr. F. Seifert[b]

[a]Institute for Communications and Information Engineering, University of Linz, Austria.
e-mail: springer@mechatronik.uni-linz.ac.at

[b]Applied Electronics Laboratory, Technical University of Vienna, Austria.

Today Surface Acoustic Wave (SAW) devices are used in many electronic products, e.g., as filters in communications applications or for signal processing in radar systems. Due to the dependence of the propagation characteristic of the microacoustic wave on many physical parameters (e.g. temperature or strain) it is possible to build SAW sensors for a great variety of physical, chemical and biological parameters. SAW sensors are not only small, rugged and show small aging rates, they also can be used as wirelessly interrogable and completely passive sensors for many hostile environments. In this contribution the SAW device principles, design procedures and technological issues as well as examples of their use in certain measurement and identification applications are reviewed.

1. INTRODUCTION

Surface Acoustic Wave (SAW) devices are made of a piezoelectric substrate with metallic structures (interdigital transducers, reflection coupler gratings) on the plain-polished surface. An electric input signal fed to a so-called interdigital transducer (IDT) is transformed by the piezoelectric effect into a microacoustic wave propagating on the surface of the SAW device. Vice versa, the SAW wave generates an electric charge distribution on an IDT which results in an electric output signal [1]. During the last years impressive progress in the performance of SAW devices was made and several innovative applications were found. SAW devices are now deployed in a great variety of electronic products. Mobile and satellite communication systems often use high performance SAW filters and every TV-set is equipped with a SAW delay line. Pulse compression SAW filters are utilised in radar systems and oscillators make use of SAW resonators. A variety of sensors for many physical, chemical and biological quantities are based on SAW devices and passive wireless identification systems can be realised using SAW ID-tags. This contribution gives a short introduction into SAW device principles and presents applications of the devices for wireless sensing and identification purposes.

2. SAW DEVICE PRINCIPLES

Every SAW device relies on propagating or standing microacoustic waves. The most important waves today are (i) the classical Rayleigh wave, (ii) the leaky surface acoustic wave (LSAW), and (iii) the surface transverse wave (STW) [2]. As substrates for the devices usually a variety of cuts of single-crystalline quartz, lithium tantalate ($LiTaO_3$) and lithium niobate ($LiNbO_3$) are used. The propagation velocity of the microacoustic waves on these substrates are in the range of 3000 to 5000 m/s. Together with IDT structures with a strip-width in the order of a few microns down to sub micron dimensions we get typical device frequencies which cover the complete VHF/UHF (30 MHz - 3 GHz) range.

The basic mechanisms in SAW devices are transduction and reflection at the metallic electrodes (strips), usually made of aluminum. IDT's and reflectors are realized as distributed elements, i.e. they consist of an array of electrode strips as is shown in Fig. 1 for the case of an extremely wideband bandpass filter for mobile communication applications [3]. Width, spacing and loading of the electrodes determine the function and electric characteristic of the SAW device. The active IDT's rely on the piezoelectric effect to generate and detect the surface waves while the passive

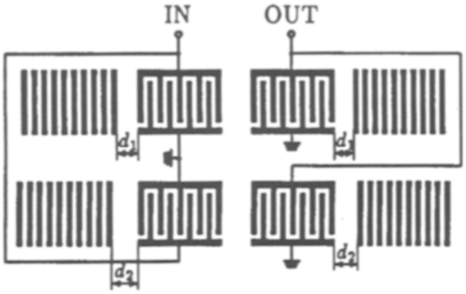


Figure 1. Schematic of interdigital transducers (IDT's) and reflector gratings

reflector gratings act as mirrors for the surface waves. The spacing and width of the electrode strips are in the order of $\lambda_s/4$ or below (λ_s : microacoustic wavelength).

Fundamental design parameters are the phase velocity v_s of the microacoustic wave and the coupling coefficient K^2, respectively. They both depend among others on the material tensors of elasticity, piezoelectricity, and dielectric permittivity, and on the mass density. As a measure of the efficiency of converting the electrical energy into mechanical energy the coupling coefficient K^2 affects the insertion loss of the SAW devices. Depending on the substrate material, the crystalline direction and the wave type, devices with different characteristics can be designed. Quartz crystal cuts operating at SAW modes exhibit a low coupling coefficient, excellent temperature stability and low bandwidths (down to 0.1%). $LiNbO_3$ on the other hand has a much higher K^2 but less temperature stability and can be used for broadband devices with relative bandwidths of up to some 10%.

The design of SAW devices is based on signal theory, network theory and field theory. A first order design is done by impulse response modeling since the impulse response $h(t)$ (and therefore the frequency response $H(f)$) of an IDT can be computed from the overlap of the electrodes as function of the position in the IDT. The field theory approach requires sophisticated numerical computations but is the most comprehensive one in which second order microwave and microacoustic effects like electrode reflections, spurious bulk acoustic wave generation, mode conversion, wave diffraction, microacoustic attenuation and dispersion, electromagnetic feedthrough between IDT's, charge distribution on the electrodes and several mechanical and electrical loading effects to name a few, can be modeled. Common field theory methods comprise full wave analysis based on partial wave analysis techniques, periodic Green's function techniques as well as finite and boundary element methods. The results of these computational extensive methods are usually fitted to analytic expressions or stored in data bases for use in the final design procedure, which is based on network theory. With this approach the SAW structure is subdivided into basic two- and three-port cells and recursively combined to yield the desired electrical behavior of the device.

SAW manufacturing has evolved along with the semiconductor industry using many techniques originally developed for the integrated circuit fabrication. SAW devices are usually produced with optical lithography and lift-off procedures to define the electrode structures. The maximum frequency of operation is determined by the minimum line width. Sub-micron processes down to 0.3 μm line width are now available, resulting in upper frequencies for today's SAW devices of around 3 to 4 GHz [4]. Production lines are able to process 4 inch wafer discs for small devices. Devices like delay lines or convolvers which process long lasting signals and are therefore longer than conventional devices (e.g. bandpass filters for communication systems) need a more specialized processing. Depending on the frequency of operation and the length of the SAW device they are mounted into TO-, dual-in-line-, SMD- or custom-designed packages.

3. WIRELESS IDENTIFICATION AND SENSOR TECHNIQUES

The main principle in SAW sensing is as follows: the transfer characteristic of the SAW device is changed by the measurand. As these transfer characteristic depends on the propagation of the microacoustic wave, many physical parameters like temperature, pressure, stress, or strain can be applied directly onto the SAW chip. Coating of the SAW surface with layers sensitive to certain chemical substances makes SAW devices useful as chemical or biological sensors. If e.g. a SAW delay line is coated with a gas sensitive layer the adsorbed gas will cause a change in the time delay. This can be translated into a frequency change if the delay line is used as a resonator for an oscillator. Of course the advantage of being sensitive to many physical and chemical quantities turns out to be a problem if, as usually is the case, only one quantity has to be measured. Therefore often the well known differential principle is used. Here two sensors are affected in opposite direction by the measurand whereas all other parameters affect both sensors in the same way. Besides their sensitivity to a great variety of physical and chemical quantities SAW sensors are small, rugged and show small aging rates.

An extremely useful feature of many SAW sensors is that they can be built as wirelessly interrogable and completely passive devices [5]. The measurement principle is based on the operation of passive identification (ID)-tags as shown in Fig. 2. The single IDT is connected to an antenna and transforms the received RF-pulse, which is typically in the range between 500 MHz to 2,5 GHz, into a microacoustic wave. Some differently spaced SAW reflectors (metal strips) are facing the IDT, reflecting parts of the wave. The train of reflected SAW pulses is transformed by the IDT back into electric pulses which are sent back by the antenna to the base station. When using the device as sensor the measurand changes the time interval between the pulses of the response signal. The difference τ_{ik} of the delays between two pulses i and k in the response signal depends on the mechanical distance L_{ik} between the corresponding reflectors and the velocity v_s of the surface acoustic wave. It is given by

$$\tau_{ik} = \frac{L_{ik}}{v_s} \tag{1}$$

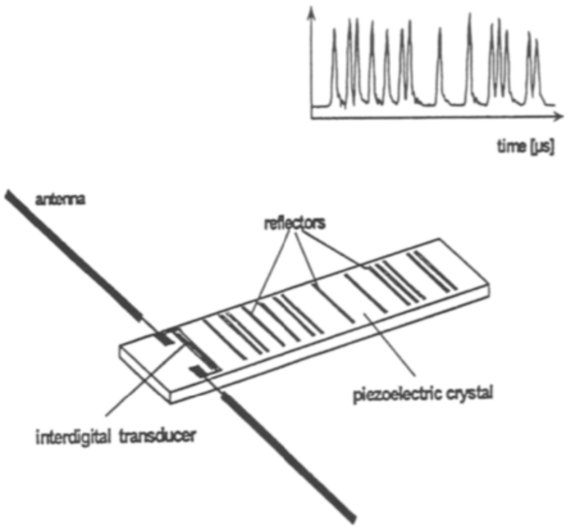


Figure 2. Wireless SAW-based sensor or identification (ID) - tag

The SAW wavelength at the frequency f is

$$\lambda_s = \frac{v_s}{f} \tag{2}$$

Using a SAW delay line type sensor, the delay τ_{ik} is evaluated from the received signal and with equation (1) the change of mechanical length or that of SAW velocity can be determined.

The response signal is delayed by the SAW travelling time which is in the order of some microseconds. Within this time interval the transmitter signal and all of its reflections from surrounding parts have decayed. So the receiver is fully sensitive to the response signal which is often more than 100 dB lower than the transmitted signal due to the insertion loss of the SAW delay line of around 50 dB and the propagation path loss. As the sensor needs no energy supply and can be read out wirelessly, after mounting it can be operated in hostile environments and even at rotating parts without any further maintenance.

As already mentioned the same device can also be used as ID-tag if the number and position of the reflectors on the substrate are used as a code unique to every device. The identification is accomplished by comparing the response signal to all known responses and finding the best match. The described ID-tags are used in a Norwegian road pricing system [6]. The interrogators are mounted above the highway tracks and the ID-tag is placed behind the car's windshield. On the SAW substrate 32 reflector positions are standardized the corresponding bits of which appears in the response signal if the position is metallized. Another ID-tag system is installed in the Munich subway system. The ID-tags are placed at the trains while the interrogation units are placed at certain points of interest in the subway system and are linked to a host computer. Despite the hostile environment due to the electromagnetic emissions from the power electronics for the drives the ID-system shows reliable operation in the 2,45 GHz ISM (Industrial Scientific, Medical) - band.

4. SPREAD SPECTRUM METHODS FOR SENSORS

The well known spread spectrum (SS) techniques used in radar systems for many years are applicable to wireless sensors, too. Advantages of SS-methods are enhanced performance in hostile electromagnetic environment and reduced power density as well as reduced peak power. For sensor applications enhanced sensitivity is an additional benefit. Chirp signals (linearly frequency modulated signals) are especially well suited for this type of sensors. The basic operational principle is sketched in Fig. 3. The chirped interrogation signal $s(t)$ (here e.g. a down-chirp) is compressed in the SAW sensor whose impulse response is a matched up-chirp. The output pulse is retransmitted to the interrogation unit. By changing the microacoustic properties of the SAW device by the physical or chemical quantity to be measured, the impulse response shifts in frequency, thereby shifting the compressed pulse in time [5].

5. SENSOR APPLICATIONS

An example for using a SAW device as temperature sensor is the monitoring of the discbrake temperature of a railroad car [6]. This is an extremely challenging task due to the high temperature differences on the discbrake as well as the hard mechanical shocks the sensor is subject to. The measurements were taken at a speed of up to 140 km/h. In Fig. 4 the temperature of the discbrake during braking for a railway station is shown. SAW sensors can be operated up to the melting point of the metallization which is 400°C for aluminum. For temperature sensors the YZ-cut of $LiNbO_3$ is favorable because of its high temperature coefficient of 94 ppm/°C.

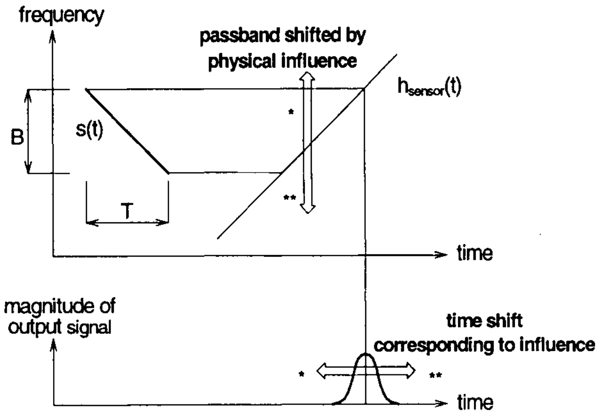

frequency
passband shifted by physical influence
$h_{sensor}(t)$
B
s(t)
T
*
**
time
magnitude of output signal
time shift corresponding to influence
*
**
time

Figure 3. Basic operational principle of a SAW chirp sensor system

A SAW pressure sensor using a Y-cut quartz membrane carrying several SAW reflectors for operation in the range of 0 to 250 kPa with a derivation from linearity of less than ±0.7% has been reported in [8] and is shown in Fig. 5. For the sensor configuration an all-quartz package was chosen and designed with a finite element method software. The hermetically closed cavity has a defined reference pressure inside and thus the sensor yields absolute pressure values. Under pressure the diaphragm bends thus changing the propagation time of the acoustic wave. By using more than one reflector and a proper combining of the reflected signals the sensors temperature dependence is almost negligible in the range from

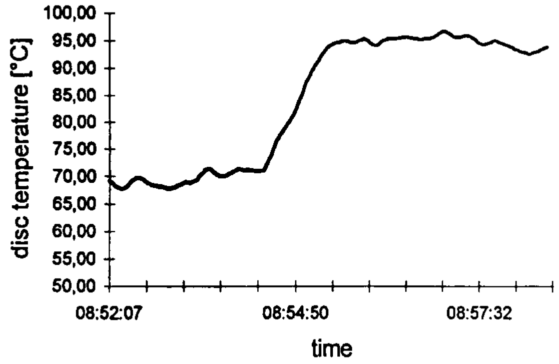


Figure 4. Temperature of a railway discbrake during braking for a station

−20°C to 100°C. A possible application for wireless pressure measurement is the monitoring of car tiers. Fig. 6 shows the measured pressure of the right front tier of a car crossing two railroad tracks and a gully [6]. Note that for this measurement an other type of SAW pressure sensor than previousley described has been used.

The measurement of torque on rotating shafts is an especially well suited application for wireless sensors. Fig. 7 shows a SAW remote sensor mounted on a shaft. With two SAW reflective delay lines in a common housing with a relative angular displacement of 90° a resolution of up to 0.2% of the nominal torque value of $M_n = 50$ Nm has been achieved over the range of $\pm 1,6 M_n$ [9]. As a substrate material $LiNbO_3$ with its high coupling coefficient has been chosen. Furthermore, due to its linear temperature response, a simultaneous measurement of temperature is possible. By measuring sum and difference of the reflected RF signal phases both sensor informations are available. The phase sum varies linearly with temperature in the range of 35°C to 70°C if the torque is held constant while the phase difference undergoes only minor changes (corresponding to 1% change of nominal torque). On the contrary the applied torque changes the phase difference but not the phase sum.

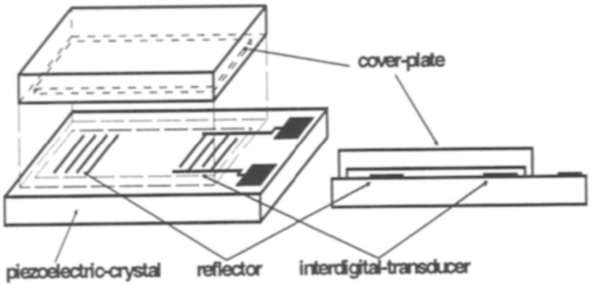


Figure 5. Schematic of a SAW pressure sensor

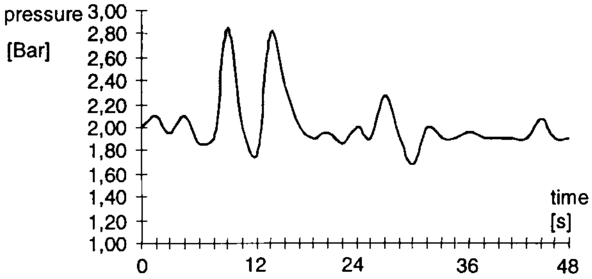


Figure 6. Pressure of the right front tier of a car crossing two railroad tracks and a gully

Another possible application for wireless SAW sensors is the measurement of current. As can be seen in Fig. 8 a classical sensor such as a magnetoresistor is combined with a SAW device. Instead of using SAW reflectors, two IDT's are used in combination with the magnetoresistors. Therefore a change in the magnetic field causes a change in the reflectivity of the IDT's. Again the differential setup is used for reducing the cross sensitivity. For currents in the range of ± 800 A a resolution of about 5% has been achieved.

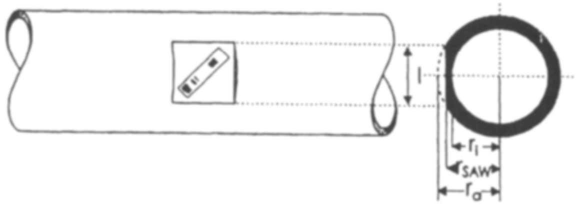


Figure 7. Principal schematic of a SAW torque sensor mounted on a shaft

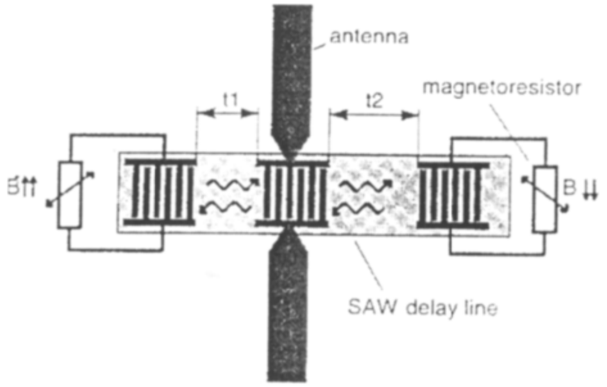


Figure 8. Wireless current sensor using a SAW device

6. CONCLUSION

An introduction into SAW device principles has been given. The underlying physical principles, design procedures and technological issues have been described. Their application as sensors and ID-tags has been reviewed. The possibility of building wirelessly interrogable sensors makes them especially attractive for measurements in hostile environments and on rotating or moving parts. A great variety of physical as well as chemical or biological quantities can be measured by SAW sensors. Examples for sensing temperature, pressure, torque and current have been given.

REFERENCES

1. D.P. Morgan, Surface-Wave Devices for Signal Processing, Elsevier, Amsterdam, (1991)
2. F. Seifert, R. Weigel, Proc. of the 27th European Microwave Conference (1997) 1323
3. G. M"uller, J. Machui, L. Reindl, R. Weigel, P. Russer, IEEE Trans. on Microwave Theory and Techniques, Vol. 41, No. 12 (1993) 2147
4. M. Hikita, N. Shibagaki, A. Isobe, K. Asai, K. Sakiyama, Proc. IEEE MTT-S (1997) 173
5. A. Pohl, G. Ostermayer, L. Reindl, F. Seifert, Proc. 1996 IEEE Inter. Symp. on Spread Spectrum Techniques and Applications (1996) 730
6. A. Pohl, F. Seifert, IEEE Trans. on Instrumentation and Measurement, vol. 46 (1997) to be published
7. G. Ostermayer, A. Pohl, L. Reindl, F. Seifert, Conf. Proc. SENSOR'97 Nrnberg (1997) 241
8. H. Scherr, G. Scholl, F. Seifert, R. Weigel, Proc. 1996 IEEE Ultrasonics Symp. (1996) 347
9. G. Scholl, T. Ostertag, L. Reindl, H. Scherr, O. Sczesny, U. Wolff, Proc. IEEE Intern. Workshop on Commercial Radio Sensors and Communication Techniques (1997) 51

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Ka-Band Doppler-Sensor

C.G. Diskus[a], A. Stelzer[a], K. Lübke[a], A.L. Springer[b], H.W. Thim[a]

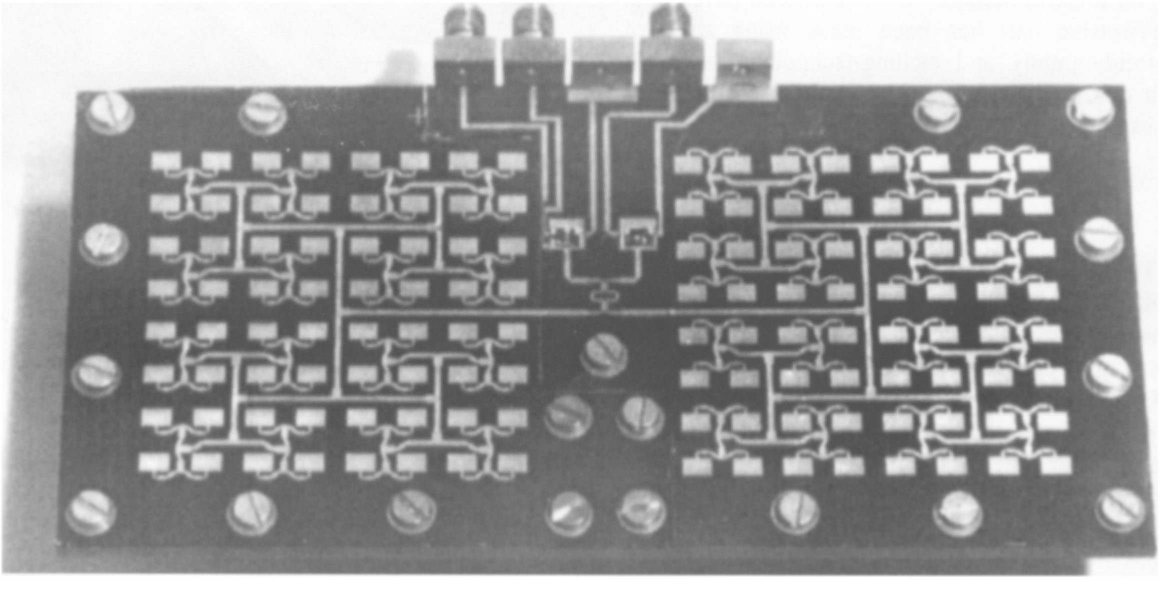
[a]Microelectronics Institute, Johannes Kepler University Linz
A-4040 Linz, Altenberger Straße 69, Austria

[b]Institute for Communications and Information Engineering, Johannes Kepler University Linz
A-4040 Linz, Altenberger Straße 69, Austria

Experimental results obtained with a 35 GHz Doppler radar unit are reported in this contribution. The measurement setup consists of a RF-front-end, signal conditioning, signal processing and a display unit. The system has been evaluated by means of a test bench, consisting of a stepper motor driven reflector and a personal computer.

1. INTRODUCTION

Millimeter-wave radar sensor systems especially suited for industrial applications were intensively developed during the last few years. Combination of a low-cost technology together with advanced electrical characteristics and high reliability is a cornerstone requirement for applications in the field of robotics and automatic control. Since the microwave front-end unit is the key unit determining the system's performance and cost, it is the most crucial part of the whole system. A 35 GHz low-cost RF front-end Doppler unit suited for industrial applications has been constructed. The unit consists of both receiving and transmitting corporate-fed microstrip patch array antennas, a microstrip directional coupler, a monolithic GaAs FECTED oscillator and an integrated single Schottky diode



mixer. Inexpensive microstrip technology has been used which yields a good compromise between cost factor and technical performance.

2. RF-UNIT

In order to minimize costs a special planar Gunn-diode (FECTED, Field Effect Controlled Transferred Electron Device) has been used instead of a sophisticated transistor oscillator. The output power of the oscillator is fed into a transmitting antenna. At the mixer diode the reflected signal from a receiving antenna is combined with the LO signal provided by a 10 dB directional coupler as shown in Figure 2.

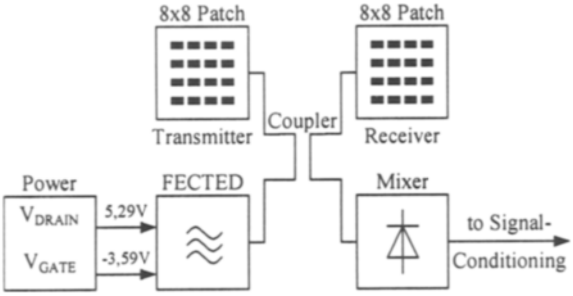


Figure 2. Block Diagramm

To achieve mixing with high sensitivity at zero bias an InGaAs-Schottky-diode has been developed. The passive part has been made using standard photolithography and etching-technology, and has been covered by a gold film in order to minimize losses.

2.1 Oscillator

The key element in a radar front-end is the signal source capable of delivering a sufficiently large signal power with high spectral purity. Although spectacular results have recently been obtained with PHEMTs [1] we have used a technologically much simpler and thus cheaper device developed in our laboratory. A schematic cross-sectional view of this device, called "FECTED" or Field Effect Controlled Transferred Electron Device [2], is shown in Figure 3. This approach relaxes the demands on the resolution of the lithography.

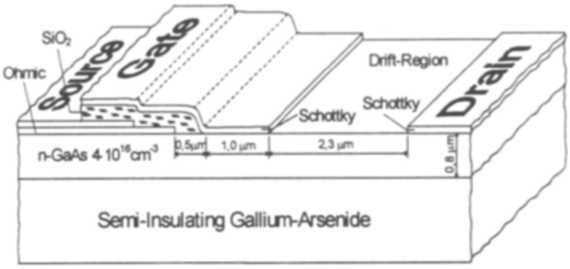


Figure 3. Cross Sectional View of FECTED

The device under question is a transferred electron device with a specially designed cathode contact, which suppresses the formation of travelling domains. This is accomplished by negatively biasing a Schottky gate contact which overlaps the cathode contact. The electron injection into the drift zone is controlled by the electric field underneath the gate, thus producing a stationary high field domain. The SEM-picture in Figure 4 shows the active zone of the device and the edge of the mesa-etched region. Further details of this FECTED have been published in [3].

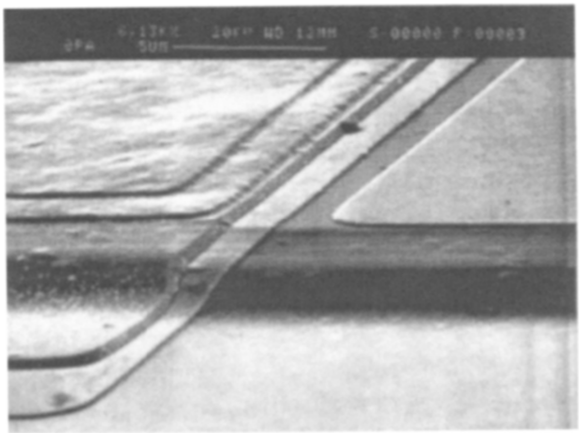


Figure 4. Active Region of FECTED

2.2 Antenna

Depending on a specific application, various requirements can be formulated for the transmitting and receiving antennas. Two linearly-polarized corporate-fed microstrip patch array antennas have been designed for the system. The first one is a 4x4 array antenna with 16 degrees beamwidth and 17.5 dB (isotropic) power gain. The second one is a 8x8 array antenna with 8 degrees beamwidth and 21.5dB (isotropic) power gain.

A relatively inexpensive RT/duroid® 5880 (a trademark of Rogers Corp.) substrate material of 0.254 mm thickness has been used for the fabrication of the antennas and the 10 dB directional coupler. More detailed description of the antennas is given in [4]. Figure 5 shows the radiation pattern of the 8x8 patch antenna.

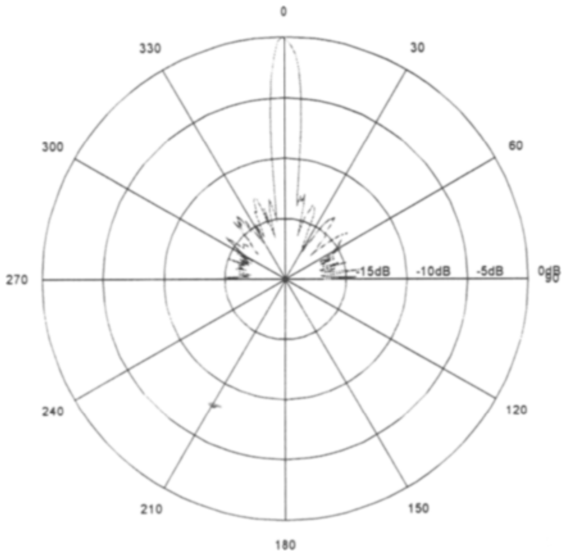


Figure 5. Radiation Pattern of 8x8 Array Antenna

2.3 Mixer-Diode

In the field of high frequency devices mostly Schottky-barrier diodes are used because of their high switching speed which results from the unipolar conduction mechanism. III-V semiconductors are preferred because of their higher electron mobility in low fields compared to Silicon. Important characteristics of Schottky-barrier detector diodes are series resistance, barrier height and junction capacitance. There are two types of detector diodes: zero-bias devices and detector diodes used with bias, which need a more complicated curcuitry. For the described radar front-end a zero bias detector diode for 35 GHz has been developed.

The voltage sensitivity β_v of a detector diode describes the output voltage supplied by the diode into the load circuit divided by the power absorbed in the diode. The goal of the optimization procedure is to make this sensitivity as large as possible [5]. The calculated dependance of β_v from the reverse saturation current I_s is shown in Figure 6 for a typical set of parameters for 10x10 μm^2 Schottky barrier diodes.

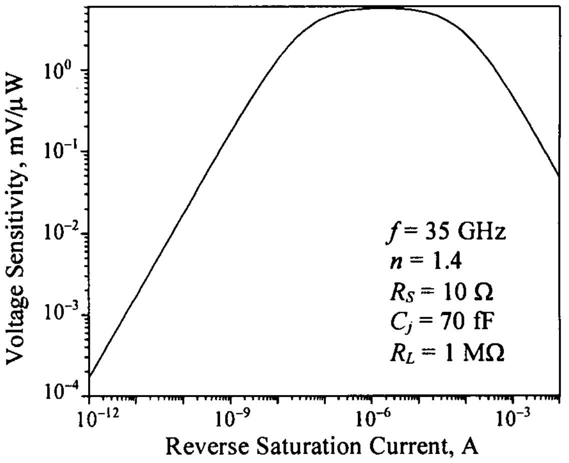


Figure 6. Dependance of β_v from I_s at 35 GHz

If the detector should be used without bias current, which simplifies circuitry, then the saturation current has to be in the range of 10^{-6} A. This corresponds to a barrier height of approximately 0.22–0.25 eV. Due to Fermi-level pinning it is hardly possible to realize such a low barrier height on GaAs. In fact the barrier height of GaAs is nearly independent of the contact metal [6]. A lower barrier height can be obtained by using $In_xGa_{1-x}As$. With increasing x the energy gap of the semiconductor decreases from 1.42 eV (x = 0; i.e. GaAs) to 0.33 eV for x = 1 (InAs). With decreasing energy gap the barrier height of the Schottky contact is also reduced. With $In_{0.38}Ga_{0.62}As$ the desired barrier height can be reached.

The diodes have been fabricated in our laboratory. Details of the structure and the fabrication process can be found in [7]. Figure 7 shows a SEM picture of the diode.

The dc parameters were determined from computer controlled *I-V* measurements. The measured forward current-voltage relationship of a typical device ist shown on a semilogarithmic scale in Figure 8.

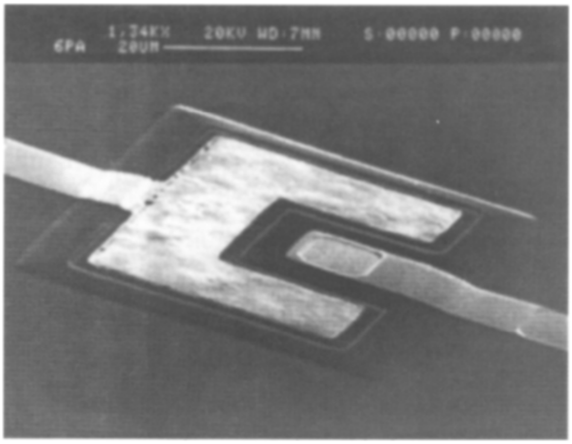


Figure 7. Mixer Diode

The semi-logarithmic plot also allows to calculate ideality factor, series resistance and barrier height of the devices. A straight line was fitted to the semi-logarithmic data to extract the parameters. Typical values for reverse saturation current I_S, series resistance R_S and ideality factor n of diodes with contact area 3x3 μm^2 are as follows:

ideality factor	1.45
saturation current	$8.2 \cdot 10^{-6}$ A
series resistance	9.7 Ω
barrier height	0.22 V
capacitance	10.7 fF

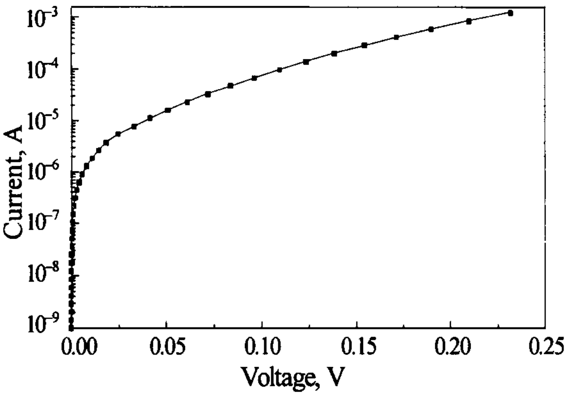


Figure 8. Semi-logarithmic plot of measured current voltage relation

From the measured junction capacitance of a diode with 100x100 μm^2 Schottky contact area, the junction capacitance for diodes with 3x3 μm^2 anode area can be calculated. The resulting cut-off frequency $f_{co} = 1/2\pi R_S C$ for these devices is approximately 1500 GHz.

A connection of the diode to a 50 Ω stripline with no further impedance matching yielded a sensitivity of 1 mV/μW. This value has further been improved by a narrowband matching circuit consisting of a T-structure integrated on the chip. With this approach the sensitivity exceeded 5 mV/μW [8].

2.4 SIGNAL CONDITIONING AND SIGNAL PROCESSING

A block diagramm of the data-aquisition and data-processing unit is shown in Figure 9. The Doppler signal from the RF-unit has to be conditioned for digitizing and digital signal processing. First, any existing DC-offset and low frequency drifts are filtered out by a high-pass input filter. Next, the signal is preamplified by a factor of 10. The second amplifier stage is combined with a low-pass filter and has adjustable gain to set the level suitable to the next stage.

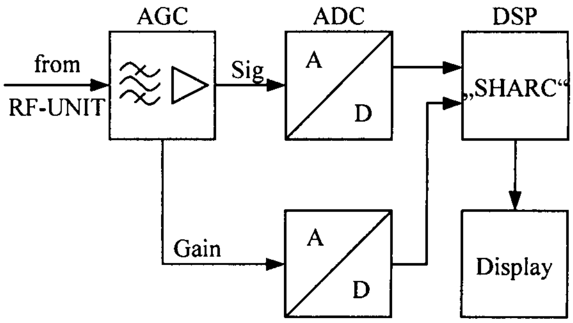


Figure 9. Block Diagramm of Data Aquisition

This is an AGC (automatic gain control) amplifier capable of handling an amplification in the range from +40 dB to –40 dB. The AGC amplifier delivers a constant output voltage level independent of the input signal thereby allowing the use of the full input range of the following analog to digital converter. The amplification of the AGC is controlled by the level of the output signal. A precision rectifier followed by a low pass and a comparator determine whether the amplification is increased or decreased.

Due to the high level of amplification in the absence of an input Doppler signal noise is also amplified to full range output. Therefore, to enhance object detection it is necessary to supply the signal processing unit with an additional information about the actual amplification of the AGC. This is performed by a second channel that measures the actual gain. The amplified and band-pass limited signal and the actual gain are digitized with a 16-bit analog to digital converter. The evaluation of the Doppler signal and the calculation of the object speed ist performed on a DSP-Board (Digital Signal Processor).

The signal processing is performed on the DSP-board equipped with a SHARC (Super Harvard Architecture Computer, Analog Devices Inc.) signal processor. Its task is to determine the frequency of the Doppler signal, i.e. the target velocity in the presence of spurios responses generated by other targets and noise from the oscillator, mixer and amplifier. The standard algorithm for spectral analysis is the FFT (fast fourier transform). To improve measurement sensitivity a calibration procedure is implemented. Without any moving target the AGC is set to maximum amplification and the system noise is measured. The spectral noise density is stored for later correction.

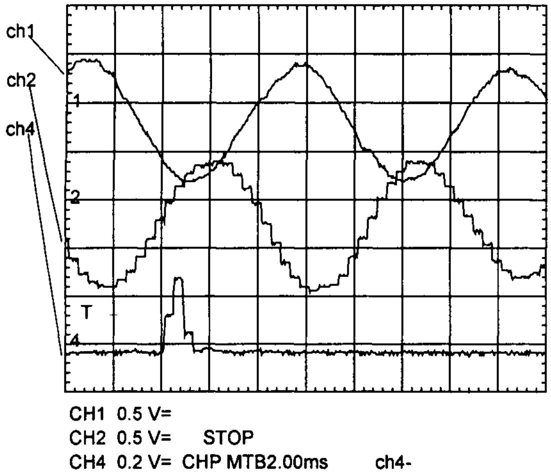


Figure 10. Measured-, Digitized-, FFT-Signal Target Velocity 0.5 m/s, Distance 2 m

As soon as a moving target is present the level of amplification is reduced in order to avoid overflow of the converter. Noise cancelation is accomplished by subtracting the weighted noise signal from the measured signal in the frequency domain. The appropriate weighting factor is derived from the value of the actual amplification factor of the AGC. In the Figures 10, 11 and 12 the amplified (ch1), digitized (ch2) doppler signals and the FFT-transformation (ch3) are shown. The phase shift between channel 1 and channel 2 is due to the program organisation and is not relevant to the calculation.

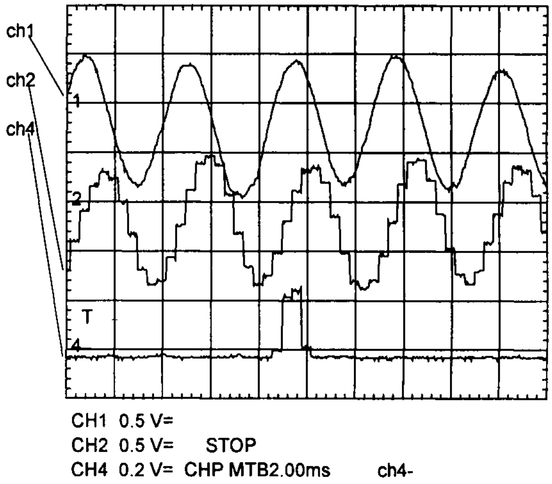


Figure 11. Measured-, Digitized-, FFT-Signal Target Velocity 1 m/s, Distance 2 m

As can be seen in Figure 12 the FFT-transformed signal is slightly different from zero because of changing system noise and drifts.

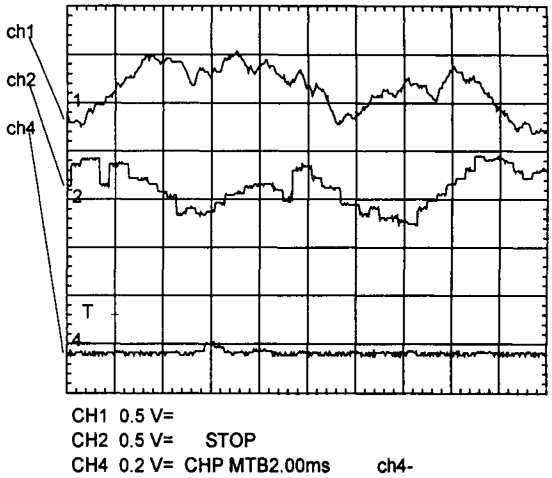


Figure 12. Measured-, Digitized-, FFT-Signal System Noise Level

The decision whether or not a moving target is present in the antenna beam is based on the level of the FFT-transformed signal. If a certain limit is exceeded a speed calculation is started. Due to the low velocity resolution of single spectral lines a weighted averaging of at least three spectral lines is performed.

3. Test BENCH

To evaluate system performance an automatic test bench has been designed. A corner reflector is moved along a precision rail towards the radar-sensor. Speed and position of the artificial target can be controlled with high accuracy and are used as reference to the measured speed.

TECHNICAL Data

Oscillator	
Type:	FECTED
Frequency:	35 GHz
Tuning Range:	350 MHz
Tuning Sensitivity:	0,6 MHz/mV
Output Power:	10 dBm
Phase Noise:	–90 dBc/Hz @ 1MHz off carrier
Detector	
Type:	Schottky
Material:	$In_{0,38}Ga_{0,62}As$
Area:	3 x 3 μm^2
Matching (50 Ω):	Microstrip T-Structure
Sensitivity:	6 mV/μW
Antenna:	
Type:	Linearly Polarized Corporate-Fed Microstrip Patch Array Antenna
No. of Elements:	8 x 8 = 64
Substrate:	RT/duroid® 5880, $\varepsilon_r = 2{,}2$
Gain:	22 dB Isotropic Power Gain
3 dB-Beamwidth	± 5°
Data Aquisition:	
AGC-Amplifier:	20–75 dB
Bandwidth:	10 Hz–3,3 kHz
A/D-Converter:	16 bit, 100 kHz
DSP:	Analog Device ADSP-2106x, SHARC
Algorithms:	FFT, Averaging

4. RESULTS AND OUTLOOK

The distance sensor has been evaluated using the above mentioned test bench. An accuracy of 0.2% has been achieved in the case of a corner reflector moving at constant speed. The maximum operational range exceeds 25 m. The amplitude of the output signal of the mixer was 0.3 mV_{pp} at this distance.
The described approach makes it possible to achieve acceptable system performance with low-cost technology. A further cost reduction and miniaturization can be achieved by monolithically integrating the system on a single chip.
An additional feature of the used FECTED oscillator is the possibility to tune the frequency by variing its gate voltage. By employing this feature the described Doppler-sensor can easily be converted to a FM-CW distance sensor.

REFERENCES

[1] Bangert, A., Schlechtweg, M., et. al., „Monolithic integrated 75 GHz oscillator with high output power using a pseudomorphic HFET“, *IEEE MTT-S*, 1994, pp. 135-138

[2] H. Scheiber, K. Lübke, et. al., „MIMIC-Compatible GaAs and InP Field Effect Controlled Transferred Electron (FECTED) Oscillators“, *IEEE*, Vol. MTT-37, No. 12, December 1989, pp. 2093 - 2098

[3] Diskus, C. G., Lübke, K., et. al., „Abstimmbarer GaAs Oszillator MMIC“, *MIOP '93*, Sindelfingen, p. 151 - 155

[4] A. A. Efanov and H. W. Thim, „Corporate-Fed 2x2 Planar Microstrip Patch Sub-Array for the 35 GHz Band“, *IEEE Antennas and Propagation Magazine*, vol. 37, no. 5, pp. 49–51, Oct. 1995

[5] Bahl I., Bartia P.; „Microwave Solid State Curcuit Design“; Wiley, New York 1988

[6] Sze S. M., „Physics of Semiconductor Devices“; Wiley, New York 1981

[7] K. Lübke, T. Hilgarth, et. al., „Zero-bias Detection with $In_{0.38}Ga_{0.62}As$ Schottky Barrier Diodes“, WOCSDICE '97, Scheveningen, The Netherlands, pp. 114–115

[8] C. G. Diskus, A. Stelzer, et. al., „InGaAs Detector Diode for Ka-Band Radar Front-End“, WOCSDICE '98, Zeuthen/Berlin, Germany

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Measuring mechanical parameters utilizing wireless passive sensors

Alfred Pohl [1], Reinhard Steindl [1], Leonhard Reindl [2] and Franz Seifert [1]

1) University of Technology Vienna, Applied Electronics Lab., Gusshausstrasse 27, A-1040 Vienna, Austria
2) Siemens Corporate Research, Otto-Hahn Ring 6, D-81739 Munich, Germany

This paper presents wirelessly interrogable surface acoustic wave sensors for applications in mechanical and automotive engineering. This type of sensors are totally passive devices, consisting of a piezoelectric substrate with metallic structures on the surface connected to an antenna. They contain neither batteries nor semiconductors or capacitors. Therefore the limits for operation are shifted to much higher parameters compared to active or semi-active devices.
Our overview focuses to SAW sensors capable to measure temperature and mechanical parameters like stress and strain applied to measurements of bending, pressure and acceleration.
The principle is discussed, setup and results of some measurements are presented.

1. INTRODUCTION

For a secure and reliable operation of today's mechanical systems more and more electronic control units have been established. This circuits are employed in power plants as well as in small motors for a wide area of applications like propulsion, etc. To supply the control systems with the information about the state of the machinery, electronic sensors are required. Therefore, a lot of devices have been invented in the last decades.
For some purposes a wireless readout is necessary. Therefore radio sensors have been developed. A radio sensor system consists of an interrogator unit and one ore a couple of sensor units. The interrogator transmits a sensor readout request signal via electromagnetic transmission or magnetic coupling, the addressed sensor unit responds with its sensor information.
Most of them are battery operated, active semiconductor circuits, others are powered by an inductive link or by a strong radio frequency (RF) carrier. The alternate voltage received is rectified and used for power supply. Many types have been published up to now [1]. The advantage of such remote sensor units is that they contain some intelligence, the measurand is collected and preprocessed. The signal retransmitted to the interrogator usually consists of a address label and the digitally coded measurand. Further some effort for coding and error correction could be implemented. One disadvantage of this type of sensors is the power required for supplying the sensor unit. Another is that the sensor units are made of silicon and moreover that they include capacitors or batteries. The temperature range for application is up to approx. 175 °C, since devices like capacitors are subject of aging, the lifetime is limited too..
Here we present another type of radio requestable sensors, based on surface acoustic wave (SAW) devices. In the next chapter, the SAW sensors and the hardware for radio request are discussed. The properties, advantages and disadvantages of these sensor systems are be pointed out. Then, some applications in mechanical engineering are presented. Finally, a brief conclusion summarizes the content of the paper.

2. SAW SENSORS

2.1 SAW delay lines

Surface acoustic wave (SAW) devices have been invented about thirty years ago. Utilizing the piezoelectric and the inverse effect, acoustic waves can be excited on the plain polished surface of a piezoelectric substrate [2]. The excitation of the SAWs and the re-conversion to an electric signal is done by metallic structures on the surface, so called interdigital transducers (IDTs). If two IDTs are applied on a substrate's surface, an electric signal is delayed in this SAW delay line device shown in figure 1.

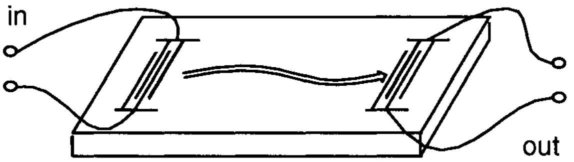


Figure 1: SAW delay line

If the substrate is affected by some physical quantity, the substrate's length and the SAW's velocity are changed.

The delay between input and output signal is the ratio between mechanical length and propagation velocity, it can be measured easily with a phase measurement between the input and the output RF signals. Another method of delay measurement is a simple impulse measurement, where the time interval between an impulse applied to the input and the response from the output is measured. Implementing such SAW devices in a feedback circuit, an oscillator can be built. The variation of frequency is a measure of the delay variation and the affect of the measurand to the substrate, respectively.

If the output transducer drawn in figure 1 is made to be reflective, a one port delay line is created, which responds to an exciting signal with the same signal attenuated and delayed.

Applying one of these principles mentioned above, a lot of sensors for thermal, mechanical and chemical measurands have been invented and published up to now [3].

2.2. Radio requestable sensors

In some cases, the location of the sensor device itself cannot be connected by wires to the measurement system. This is true, if the sensor has to be located on the surface of a fast rotating transmission axle, on a bucket wheel of a turbine, on a disc brake or within a car tire. Further it could have to be placed within a chamber made of concrete enclosing materials emitting hard radiation or in an area filled with poisonous gas. In these and many other cases a radio link has to be installed. The sensor unit itself must be supplied with energy and the measurand has to be read out by the distant sensing system wirelessly. If the remote sensor is interrogated, to avoid masking of the response, the interrogation signal has to be separated from the sensor's response in time or frequency, respectively.

For separation in time, the delay of SAW devices is useful. To get a wirelessly requestable device, the reflective devices discussed above only have to be connected to an antenna [4, 5]. The according structure and the signals for interrogation are sketched in figure 2. The radio request unit transmits an interrogation signal. Received by the sensor's antenna, the signal excites an SAW on the substrate's surface.

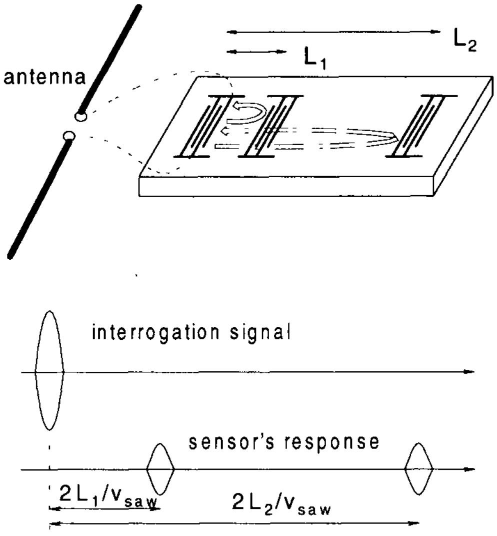


Fig. 2: Wirelessly requestable passive SAW sensor and signals for radio request

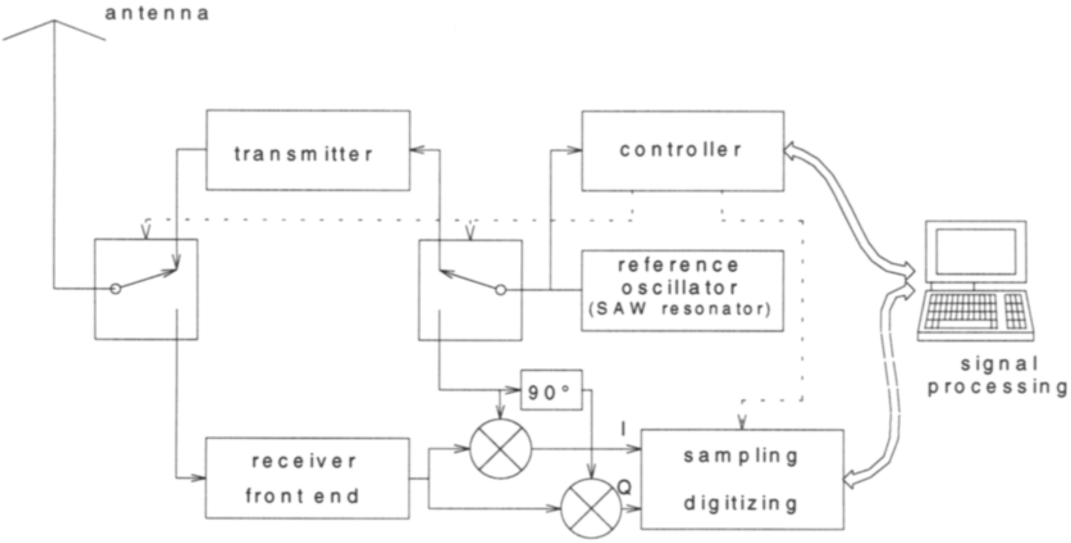


Fig. 3: Hardware of the radio request unit for passive SAW sensors

After the delay time interval of the reflective propagation path, the device's response is retransmitted by the sensor antenna.

The response signal is separated in time from the interrogating signal. Since the delay time is about some microseconds, all other echoes from electromagnetic reflections have been faded away and the sensor's response can be detected exclusively.

In figure 2 two reflectors are sketched. Interrogating with an RF burst, this device responds with two bursts with a differential delay of $\Delta\tau = (L_2-L_1)/v_{saw}$. The delay $\Delta\tau$ only depends on the length of the substrate and the SAW's velocity, the sensor and the affect to itself, influences originating from the wireless transmission path, the movement of the sensor, etc. are compensated.

The sensor itself is a totally passive device, the energy for retransmission of the response is gained from the RF request signal.

Like discussed above, the measurand affects the sensor and changes length and SAW's velocity by changing the crystal constants of the piezoelectric substrate, respectively. For measurement, temperature, mechanical stress or strain can be applied to the sensor, it can loaded to be bent or it can affected by mass loading by an adsorbing coating to become sensitive for gases to observe.

If the sensor is affected, the sensor's response to the interrogation signal is scaled in time by a factor $s=1+\varepsilon$, all delays τ_i become $\tau_i'=s\cdot\tau_i$.

The signal processing has to evaluate this scaling from the time intervals between the response burst signals [6].

2.3 Radio request unit and signal processing

The electronic hardware environment for the radio request unit for such SAW delay line sensors consists of a transmitter, a receiver, a controller and a signal processing unit, calculating the measurand's quantity from the differential delay or the sensor's response scaling, respectively. In figure 3 the block diagram of an interrogation system is drawn.

The transmitter section generates and transmits the RF radio request signal (usually a burst). The sensor's response signal is amplified and detected coherently, utilizing a quadrature demodulator. The digital signal processing unit evaluates the differential delay $\Delta\tau$ and calculates the measurand's magnitude.

If only the delay between the envelope of the response bursts is measured, the resolution will be poor. A 10 MHz RF system's bandwidth, yields a duration of about 100 ns of the interrogation burst and of the impulses retransmitted from the sensor.

The sensitivity shall be derived for temperature measurements, the sensitivities for other measurands can be calculated analogously. Looking at the envelope of the response in time domain only, shifts of about 10 ns can be detected. For example for temperature measurements a Lithiumniobate ($LiNbO_3$) delay line with a delay sensitivity of approx. 80 ppm can be applied advantageously. For a sensor with an absolute delay of 6 μs (20 mm propagation distance, bidirectional), a delay of 10 ns represents 1660 ppm of the full span. Therefore the resolution is approx. 1660/80 ≈ 20 K, rather poor for precision measurements.

If the coherent receiver concept shown in fig. 3 is applied, the reference oscillator in the system, the transmitted burst was derived from, can be used as ruler in time, the phase of the response impulses can be measured.

The phase measurement's resolution depend on the noise of the received signal and the digitization of the baseband quadrature signals. If it is processed using 8 bit in parallel and the dynamic range is modulated fully, the phase can be measured with a resolution of better than one degree. One degree in time domain represents a delay of 1/360 of one RF period. For a 433 MHz interrogator, one degree corresponds to a delay of 2.3 ns / 360 = 6.4 ps. So, for the assumed parameters, the resolution can be enhanced from 10 ns to 6.4 ps or from 20 K to 0.0128 K in optimum.

In practical use the resolution for temperature measurements is limited to approx. 0.1 K due to noise and insufficient modulation and level control, respectively, at the sample unit.

2.4 Sensor properties

In comparison with other wireless sensor systems the main advantage is, that SAW sensors are total passive. Since they contain no semiconductors nor capacitors or other electronic components, they are not limited in such a narrow manner by temperature, radiation, etc. The upper temperature limit is given by the melting temperature of the used metals and the piezoelectric curie temperature of the substrate (here the piezoelectric effect vanishes), respectively. The upper limit for materials usual used is about several hundreds degree centigrades. SAW sensors operating at approx. 1000 °C have been published [7]. SAW sensors only consist of a substrate (e.g. Quartz, Lithiumniobate, etc.) and planar metallic structures on its surface. The capability to withstand mechanical load is that of the crystal itself.

The problems in practical use are, that the sensor have to be encapsulated hermetically, since every dust on the surface will cause absorption of the SAW and cause a fail of the sensor. Further, in comparison with radio sensors powered by a battery, similar to a RADAR scenario with passive reflectors, the attenuation of the electromagnetic wave between the interrogator and the sensor is with more than the fourth power of the distance. So, the feasible distance for interrogation with permissible radio transmission is limited to be a few up to 20 meters. Usually, like in RADAR technique, coherent integration measures are applied, the results are averaged in time for a number of interrogation cycles. This yields more signal's energy E_s in the total sensor's response used for further signal processing and reduces the errors due to noise while calculating the measurand. Further it helps to match to the governmental regulations for wireless telemetry systems. The energy E_S is divided to k interrogation signals, the power required for each of these is by a factor k smaller .

3. APPLICATIONS

For measurement of temperature, e.g. at brake discs, and for identification purposes, a lot of publications have been presented in the recent years [8, 9, 10].

In this chapter, to prove the versatility, some examples for the implementation of a mechanically loaded sensor for force and displacement measurements are shown. Many other mechanical parameters like torque, pressure and acceleration can be attributed to be loaded in this way.

3.1 Stress and strain

If a SAW substrate is stress loaded or is exposed to a mechanical strain (stretching or compressing), the elasticity constants and the velocity of the surface acoustic wave, respectively, are changed. This can be detected by a time or phase difference measurement, processing two or more burst (bits) of the sensor's response in a differential mode.

Figure 4 points out the phase difference between two bursts of the response versus the relative extension.

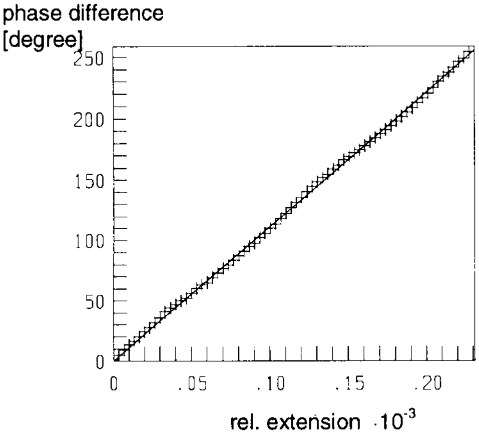


Fig. 4 Sensitivity of a strain sensor

This principle is applied e.g. for torque measurements [10], where two sensors are fixed right angled onto a shaft (fig. 5). Torque leads to a torsion of the axle and to extension of one and compression of the other sensor. The measurand can be evaluated from a differential measurement, therefore the temperature drift is compensated.

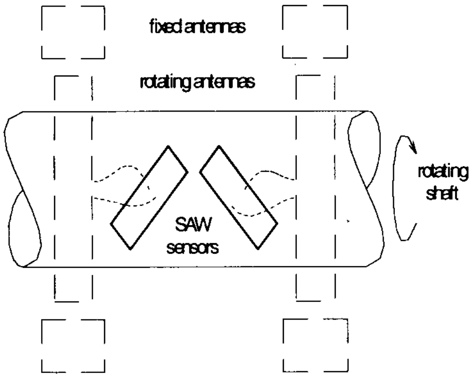


Fig. 5 Setup for radio requestable torque measurements

3.2 Measurement of displacement

Here, the SAW sensor is loaded to be bent, again the differential phase shift between reflectors is evaluated. Figure 6 shows the phase shift between two reflectors with an unscaled spacing of 3 µs versus displacement of one edge with the sensor fixed at the opposite edge.

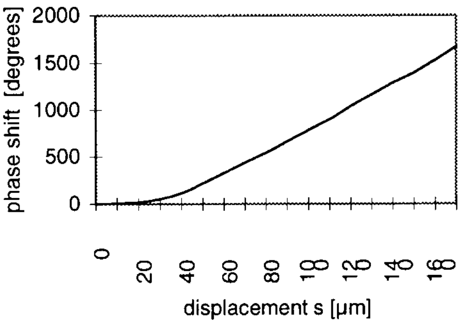


Fig. 6 Measured phase shift versus bending the SAW sensor

Based on mechanical displacement, a lot of parameters can be measured. Precise position detection with directly affected sensors is feasible. Other measurands have to be converted to a displacement: A useful converter is a spring or a membrane, converting a force or a force per area (pressure) to a linear shift.

Applying this, a lot of measurements have been done with radio requestable passive SAW sensors mentioned above. Measurements of air pressure in car tires have been published [11, 12]

In mechanical engineering vibrations and acceleration are important measurands. Since the force is equal to the product of mass and acceleration F=m·a, a acceleration can be converted to a force using a seismic mass, radio requestable SAW sensors also can be applied for wireless acceleration measurement.

To prove this, we fixed a sensor assembly to a dart arrow and observed the deceleration in the phase of invading the target. Preliminary results are shown in figure 7.

When the head arrives and invades, due to the cylindrical head, the arrow is decelerated linearly.

Then, if the handle arrived at the surface, the motion is converted to deformation.

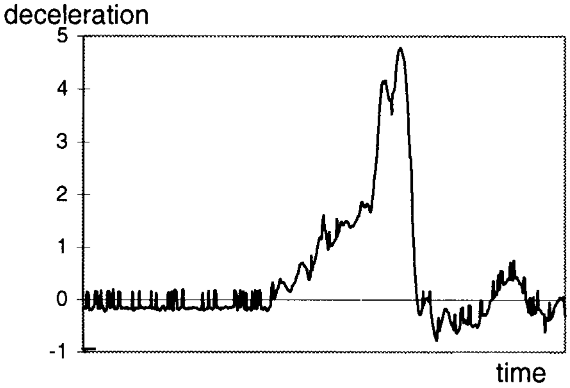


Fig. 7: Deceleration measurement of a dart arrow arriving and invading the target, measured wirelessly with a passive radio requestable SAW sensor

4. CONCLUSION

Passive radio requestable sensors using surface acoustic wave devices was introduced. The high resistance of these sensors against environmental parameters, destroying active sensor devices containing semiconductors etc. was confronted to their need for a higher effort in signal processing. Some applications for remote measurement of mechanical parameters were presented. It was shown, that in future some measurements could be performed much easier utilizing SAW sensors.

REFERENCES

[1] RFID-Handbook, ISBN 3-446-19376-6, Carl Hanser Verlag, Munich, Germany, 1998

[2] D.P. Morgan, *Surface Wave Devices for Signal Processing*, Elsevier, Amsterdam, 1985.

[3] F. Seifert, W.E. Bulst and C.C.W. Ruppel, Mechanical sensors based on surface acoustic waves, *Sensors and Actuators, A44* (1994), pp. 231-239.

[4] P.A. Nysen, H. Skeie and D. Armstrong, System for interrogating a passive transponder carrying phase-encoded information, *US Patent Nos. 4 725 841; 4 625 207; 4 625 208* (1983-1986).

[5] L. Reindl, F. Müller, C. Ruppel, W.E. Bulst and F. Seifert, Passive surface wave sensors which can be wirelessly interrogated, *International Patent Appl. WO 93/13495* (1992).

[6] A. Pohl, G. Ostermayer, L. Reindl and F. Seifert, Spread spectrum techniques for wirelessly interrogable passive SAW sensors, *Proc. IEEE International Symposium on Spread Spectrum Techniques & Applications 1996*, pp. 730-734.

[7] J. Hornsteiner, E. Born, E. Riha, Surface acoustic wave sensors for high- temperature applications, Proc. IEEE Frequ. Contr. Symp., 1998.

[8] W.E. Bulst, C.C.W. Ruppel, Developements to watch in surface acoustic wave technology", Siemens R&D Special, Spring 1994, pp. 2-6.

[9] W. Buff, F. Plath, O.Schmeckebier, M. Rusko, T. Vandahl, H. Luck, F. Möller, Remote sensor system using passive SAW sensors, Ultrasonics Symp. 1994, pp. 585-588.

[10]U. Wolff, F. Schmidt, G. Scholl, V. Magori, radio accessible SAW Sensors for Simultaneous Non-Contact Measurement of Torque and Temperature, Proc. Ultrasonics Symposium 1996

[11]H. Scherr, G. Scholl, F. Seifert, R. Weigel, Quartz pressure sensor based on SAW reflective delay lines", Proc. IEEE Ultrasonics Symp. 1996.

[12]A. Pohl, G. Ostermayer, L. Reindl, F. Seifert, Monitoring the tire pressure at cars using passive SAW sensors, Proc. IEEE Ultrasonics Symposium, Toronto, 1997.

ACKNOWLEDGMENT

The authors thank Siemens Munich and the staff of the Corporate Technology Laboratory ZT KM for supporting our activities, producing excellent SAW devices and for helpful discussions.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

A new two-axis attitude sensor based on free pendulum inclinometer

Nicola Getschko

Department of Mechanical Engineering - Polytechnic School - University of São Paulo
Department of Electrical Engineering - Catholic Pontifical University of São Paulo
Av. Prof. Mello Moraes, 2231 -São Paulo -SP - CEP.: 05508-900 - BRASIL - E-mail: ngetschk@usp.br

ABSTRACT

This paper describes the design steps and tests of a new, low cost, attitude sensor called AGS. With digital features, it was proposed to control two axis attitude (pitch and roll) of the handpiece, in clinical procedures performed by dentists. Ergonomic and miniaturization requirements, besides precision and dynamic response, were some basic parameters of the design methodology adopted. The AGS controls handpiece and tool attitude in real time, without any constrains to the dentist hands movement. To avoid the traditional magnetic or viscous procedures disadvantages, a virtual digital damping system, based on a new software algorithm, provided the required sensor damping treatment. To test the AGS in clinical conditions, a group of dentists performed an abutment preparation test. The results using AGS, compared with normal preparations procedures, have showed a clear reduction in angular deviations from the reference guiding planes.

1. INTRODUCTION

The quality of many odontological procedures, as the applied in the abutment teeth preparations for removable and fixed partial dentures, depends on a basic mechanical factor: to reproduce, in the patient teeth, the extra-oral planned preparation (Applegate, 1965; Mack, 1980). In these cases, the preparation must be made with the tool axis parallel to a spatial reference called insertion-path of the prosthesis (PP', figure 1) or guiding plane (Jochen, 1972). Solutions for this problem, proposed by many authors (Froner, 1987), have shown a little or no viability.

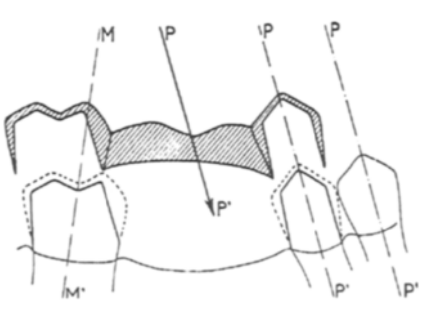


Figure 1.- Insertion-path (PP') of a partial fixed denture. - (Le Huche, 1972)

Similar problems are found in dental implants, where the implantation procedure also needs a system of adequate and reliable guidance (Branemark, 1987). Malaligned implants can compromise the prosthesis bio-mechanical and aesthetics considerations (Casellini, 1995). Most of the systems considered are based on the mechanical constrained guidance of the handpiece (Göransson, 1975; Bass, 1988; Gold, 1985).

The Attended Guidance System - AGS - looked for, instead of restricting the movements of the hand of the dentist, to minimize its interference through the simple indication of a positional deviation occurrence, and the direction to correct the error.

2. TECHNICAL REQUIREMENTS

The AGS was designed to reach, specifically, the requirements of the previously cited odontological procedures. Their definitions were based on clinical data gotten from the professors of FOUSP, bibliographical research and anthropometric and ergonomic studies (Damon, 1966; Mack, 1980; Burguess, 1986).

Some important design parameters were:

a-) Attitude control (pitch and roll) of the handpiece in real time.

b-)Minimum interference with the clinical procedures: no mechanical constraints linking the sensor to the reference; lesser possible weight; no interference in the visual field - lesser possible size; information of error by means of a sound and a visual signal.

c-)Patient Security - does not have mobile parts displayed or in the inward of the patient mouth. Use of low electric tension (5V).

d-)Comfort of the patient - minimize any possible interaction with the patient.

e-)To support asepsis by Liquid of Dakin (0.5% solution of sodium hipochlorite).

f-)Resolution better than 1°.

g-)Frequency Response higher than 2 Hz and Time Response lower than 0.5 s - It was established that will be possible to have a new event occurring to each 1s, so from the Theorem of the Sampling (Stallings, 1988), a frequency response better than 2 Hz was enough.

h-)Maximum measurement speed bigger that 480°/s (maximum angular speed reached by the toggles of the elbow and pulse).

i -)Measurement range of 50° (+ / - 25°).

j-)Capacity of simple change of the system parameters.

k-)Maintenance and electromechanical adjustments free.

3. INERTIAL SENSORS

The need of absence of mechanical constraints (parameter 2-b), led to the adoption of the inertial principles for the attitude determination. Systems based on gyroscopic sensors and inclinometers had been analyzed, being chosen the last one. Conventional gyroscopes were discarded due their high weight, size and cost not compatible with AGS. "Solid-state" gyroscopes have acceptable dimensions and costs, however they present disadvantages as low accuracy, variations of gain and noise sensitivity (Weiss, 1993; Barshan, 1995; Marselli, 1996; Scheding, 1997). The use of inclinometers for the attitude control is a good option (Kielland, P., 1995), congregating advantages like adequate accuracy, low cost and direct use of the signal, without need of integration. One important disadvantage of the existing models, for applications where the miniaturization is necessary, is the excessive size and weight (from 0.05 kg to 0.7 kg).

4. OPERATIONAL PRINCIPLES

The AGS adopted as principle of the attitude measurement, the determination of the angular position of two pendulums, oscillating according to two perpendicular axles, as it shows figure 2.

The set of pendulums was attached to the handpiece (figure 3), with the electronic circuit. This circuit was used for the optic reading of the position of each pendulum and for the coding and transmission of the data to the processing and control circuit. This subsystem was called "Attitude Measurement Device - AMD."

During the set up procedure, the AGS acquires the values of the angles αi and βi, referring to the position of the insertion-path. After this, as the dentist moves his/her hand, AGS starts to measure αc and βc, referring to the position of the handpiece, and calculates their differences from αi and βi, in real time. Since the module of the difference, between αi and αc or βi and βc, exceed the value of the defined tolerance, the system emits a sonorous and a visual alarm so that the proper corrections can be made. Using the pendulums and the not adopting of any traditional oscillations damping system (viscous or the magnetic) allowed a great reduction in the volume and in the mass of the AMD. However, to make AGS able to control, in real time, the attitude of the handpiece, it has been incorporated an innovative concept of digital virtual damping. It was carried through by means of a software algorithm - SDS -, capable to foresee the final equilibrium position of the pendulum, from the determination of

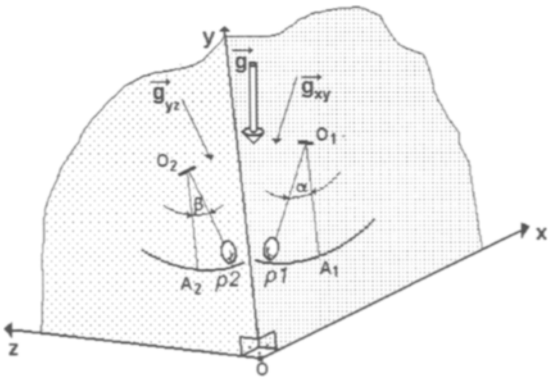


Figure 2. Operational principles of the AGS

its instantaneous angular readings (Getschko, 1997).

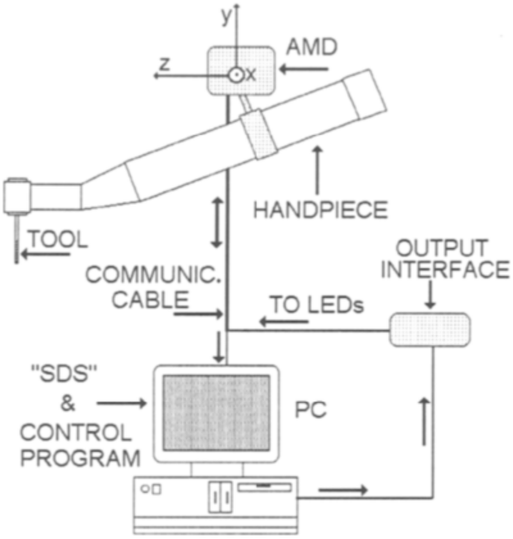


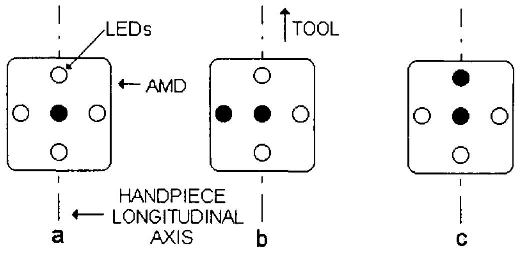
Figura 3. - Schematic view of the AGS

4.1. Operational platform

A standard IBM-PC platform was chosen to run the digital damping software - SDS - and the Control Program. This choice was due to its versatility and the great number of dentists using this type of personal computer in their offices.

4.2. Interfaces

The serial interface was used for data input , generated by the AMD, while the output of the positional correction and error signals was made using the parallel port of the PC. AGS set up and configuration were made using keyboard and monitor display. During tooth preparations procedures, the information was transmitted to the user (dentist) by two means: a) the built-in PC sound system, to alert the dentist of an attitude deviation greater than the tolerance; b) visual, using a set of five LEDs, placed on top of the AMD, inward the visual field of the dentist, to inform the occurrence of the positional error and which was the direction to correct it (figure 4). The use of two systems of error alert, was very useful to become the operation of the AGS most efficient and comfortable to the dentist. Only after the sound alert the dentist deviates, partially, his/her focus of vision from the tooth, to observe the correction indications of the set of LEDs.



a-correct position b- Turn left c- Turn forward

Figure 4. - Top view of the AMD and LEDs set

4.3. Control Program

The Control Program of the AGS, has incorporated the digital damping software - SDS - and has done the following tasks:

- read the position of the AMD;
- decide if the oscillation is symmetrical or not;
- identify an extremity of the pendulum movement;
- calculate the position of virtual equilibrium;
- acquire the set up parameters: calibration (degrees/ pixel) and tolerance.
- compare the value of the error, between the calculated position of the pendulums and the reference, with the user defined tolerance, and to emit the signals for "ok" or "error."

4.4. Pendulum Design

One of the most important concerns was to minimize their volume and mass. SDS demands, at least, a full pendulum oscillation to predict the equilibrium position, thus the physical shape of the pendulum was as a circular sector of 108° (2 x 50° + 8°- due to constructive reasons). Optical marks for the angular position reading had been printed in the pendulum periphery. A sensor (two photo-diodes + one LED) with optical pitch of 0.32 mm was chosen, and the adopted optical radius was 15.5mm. As the photo-diodes and the AMD electronic circuit provide a 2-channel quadrature output, two pulses are generated for each mark and is possible to define the direction of the movement.

Figure 5 shows a schematic draft of the pendulum built with transparent polymethylmetacrilate. Beyond the optical pendulum resolution (0,59°/pixel), it was necessary to guarantee a suitable mechanical resolution, which depends of the bearing friction.

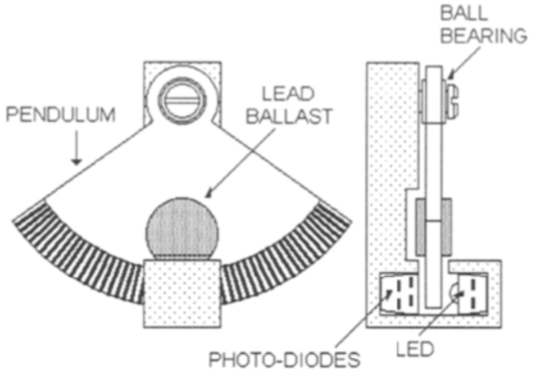


Figure 5. - AGS Pendulum

An NSK 681X miniature ball bearing was used. The influence of the mass of the pendulum ballast on its friction angle (mechanical resolution) was studied (Palmgren, 1959;Harris, 1984). For the adopted ballast mass, 0.877 grams, was found that the theoretical friction angle was inferior the 0.0025 rad (0.14 degrees). For these conditions, the estimated life of the rolling bearing was greater than 10^{12} hours.

5. PERFORMANCE TESTS

Test procedures of the AGS had been divided in two groups: general performance and clinical tests. The results of the general performance tests are shown in table 1. The clinical test had been planned with professors from the Dept. of Prosthesis of the Faculty of Odontology of the Univ. of São Paulo - FOUSP-, to assure that it would represent the actual clinical process. The *in- vitro* test, simulated the preparation of four mandibular abutment teeth for the implantation of a partial removable denture. To minimize the influence of uncontrollable factors and to allow a better control of the results, the teeth had been replaced by brass cylinders, with average compatible dimensions. Eleven mandibles were cast from a real mandible mold.

Table 1 - Technical Characteristics

Resolution - mechanical	0.15 +/- 0.01	(degrees)
Resolution - optical	0,59	(degrees)
Frequency Response	2,8 Hz	
Operational Range	+/- 25	(degrees)
Linearity	100%	(oper. range)
Repeatability	>99,3%	(static)
	>99,7%	(dinamic)
Maximum Ang. Rate	<1000 degrees/s	
AMD Dimensions	37x53x29 mm-external	
AMD mass	33 grams	

The teeth of one mandible, had been prepared using a paralelometer, to get the ideal preparation surface and to be used as a standard model. Using still the paralelometer, a cylindrical pin, whose axle coincided with the prosthesis insertion-path, was attached to the standard mandible. This way, the AGS could acquire the positional reference, independently of the spatial position of the mandible. For a real clinical use, this reference pin can be substituted by one jig imprisoned to the patient teeth, as it shows the figure 6.

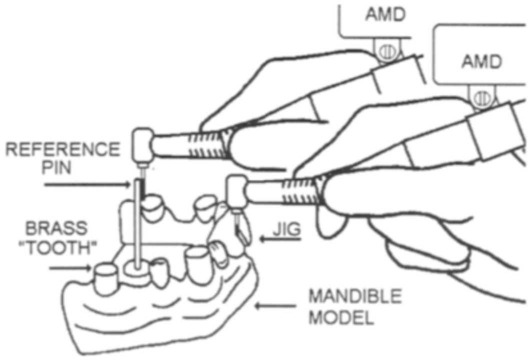


Figure 6. - Acquiring references for the AGS

Five dentists, with equivalents skills in this kind of preparation, had been assigned for the accomplishment of the tests. Each dentist received two mandibles, assembled into an anatomical dummy, and the standard model. The teeth of the first mandibles had been prepared using just the standard model as visual reference, according to traditional clinical practice. The second mandible was prepared only with aid of the AGS. Tooth preparation order was defined by the dentist, based on his/her personal preferences.

6. RESULTS

The preparation surfaces of each tooth of the mandibles and the standard model had been digitalized in one CMM Mitutoyo - model BN-710. Collected data had been processed in a CAD program, to establish a geometrical matching between the two surfaces(test and reference). The error was defined as the average angular deviation -"-between the intersections of the two surfaces with four different plans that contained the insertion-path, as shows figure 7.

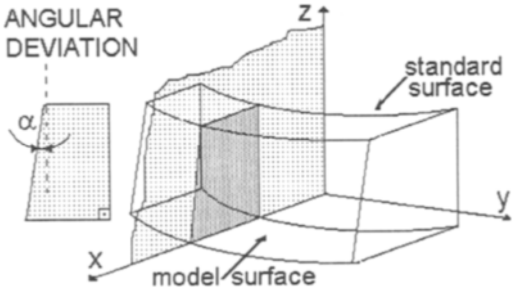


Figure 7. - Angular Deviation Measurement

6.2. Statistical Analysis

The numerical data of the angular deviation averages of each tooth, had been compared using analysis of variance and the significance tests, so that the behavior of the averages was established for each relevant factor. The results had been the following ones:

Analysis of the process - evaluated the influence of the dentist skill and the tooth location in the average error.

1 -) Without the use of the AGS - traditional method:

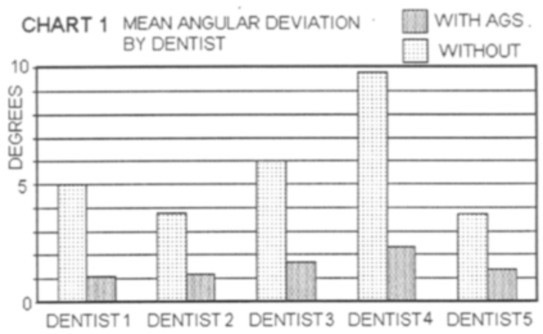
a-) There is a significant difference due to the dentists, at a 5% level, but not due to the teeth location.

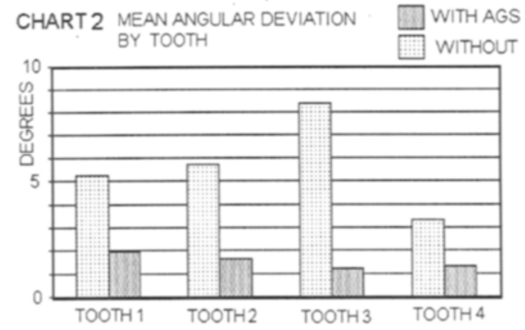
2 -) With the use of the AGS:

a -) There is not a significant difference nor due to the dentists, neither due to the location of teeth, at a 5% level.

When the angular deviation averages of the dentists had been compared, the analysis of variance showed that, for the four teeth preparation in each mandible, it had a significant difference, at a level of 5%, related to the use or not of the AGS, however it did not have difference due to the dentists.

The significance tests showed that, for each tooth, the averages of the angular deviations were lesser with the use of the AGS. Charts 1 and 2 show some results.





7. CONCLUSIONS

From the need of handpiece and tool attitude control in many odontological procedures, it was proposed, designed, built and tested a system to control attitude called AGS (Attitude Guiding System). To attend the ergonomic and technical requirements, a free-pendulum inclinometer was developed, using a new type of damping, based on a numerical algorithm, called SDS (Software Damping System). The performance tests of the AGS had shown its viability for use in other engineering applications where miniaturization will be important. The clinical tests had shown that using AGS in the teeth preparation surfaces, leads to a clear reduction in the angular deviation average from the reference standard. To extend the validity of the clinical use, new tests should have to be executed, with bigger samples and with the application of *in vivo* tests. Another application to be tested is the use of the AGS in odontology students training.

REFERENCES

APPLEGATE, O.C. Essentials of Removable Partial Denture Prosthesis, Sanders, 1965.

BARSHAN, B. et alli Evaluation of The Solid-State Gyroscope for Robotics Applications, IEEE Transactions on Instrumentation & Measurement, February 1995, Vol. 44, pp. 61-67.

BASS, E.V. et alii Controlled Tooth and Mouth Preparation for Fixed and Removable Prosthesis, J. Of Prosth. Dentistry, March, 1988, Vol. 59, No. 3, pp. 276-280.

BRANEMARK, P.I. et alii ,Prótesis Tejido-Integradas, Quintessenz-Verlag, 1987.

BURGESS, J.H. Designing for Humans: The Human Factor in Engineering, New Jersey, U.S.A., Petrocelli Books, 1986.

CASELLINI, R.C. et alli Restoring Malaligned Implants: Clinical and Technical Perspectives, CDA Journal, vol. 23, No. 3, 1995.

DAMON, A. et al, The Human Body in Equipment Design, Cambridge, USA, Harvard Univ. Press, 1966.

FRONER, E.E. Contribution to the Study of Guiding-Planes in Partial Prosthesis, master thesis, FOUSP, São Paulo - SP, Brazil, 1987.

GETSCHKO, N., Stabilization in Digital Low-Cost Piezometric Sensors, Proceedings-Abstracts of COBEM'97, pg.295 (CD-rom full version) conf. code COB969, Bauru - SP, Brazil, 1997.

GOLD, H.O. - Instrumentation for Solving Abutment Parallelism Problems In Fixed Prosthodontics, J. of Prosth. Dentistry, February 1985, Vol. 53, No. 2, pp. 172-179.

GÖRANSSON, P., The new Paralleling Instrument, Paramax II, and the Kodex Drills ," J. Of Prosth. Dentistry, July 1975, Vol. 34, pp. 31-34.

HARRIS, T. A., Rolling Bearing Analysis, John Wiley and Sons, 1984.

JOCHEN, D. G., Achieving Planned Parallel Guiding Planes for Removable Partial Dentures, Dental Technology, June1972, pp. 654-661.

KIELLAND, P., et al., Using DGPS to Measure the Heave Motion . . ., Navigation the 90's: Technology, Applications, and Policy Proceedings of the Annual Meeting, U.S.A., 1995, pp. 495-504.

LE HUCHE, R., Inlays Et Onlays - Bridges sur Dents Pulpées, Julian Prélat, <. Edition, Paris, 1972.

MACK, P.J. - The Theoretical and Clinical Investigation into the Taper Achieved on Crown and Inlay Preparations, J. of Prosth. , May 1980, vol.7, pp. 255-265.

MARSELLI, C. et alli, Error Correction Applied to Microsensors in the Navigation Application, Institute of Microtechnology, University of Neuchâtel, September 1996.

OPTOELETRONICS, Designer's Catalog, Hewlett Packard, 1995

PALMGREN, A., Ball and Roller Bearing Engineering, Philadelphia, 1959.

SCHEDING, S. et alii, Experiments in Autonomous Underground Guidance, proceedings - ICRA'97, 1997,pp. 12.

STALLINGS, W., Data and Computer Communications, MacMillan Publishing Company, USA, 1988.

WEISS, G., Usage of the Vibration Gyroscope for Orientation Estimation, Univ. of Kaiderslautern, 1993, pp. 3.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A VISION ASSISTED MECHATRONIC SYSTEM FOR ARC WELDING

Luís Gonzaga Trabasso[a] and Carlos Eduardo Ferreira Silva[b]

[a] Department of Mechanical Engineering, Technological Institute of Aeronautics, São José dos Campos - State of São Paulo – Brazil

[b]General Motors of Brazil, São José dos Campos – State of São Paulo - Brazil

This paper presents a Mechatronic Welding System (MWS) for arc welding, based upon ordinary welding mobile units added by a computer vision system which assists the MWS controller for repositioning of the welding torch as well as for seam tracking. The main computer vision algorithms are based on border detection which enables the MWS to be used in a multipass welding process.

1. INTRODUCTION

Industrial companies which use extensively the arc welding process seek to improve the productivity of the process through the use of automated systems which yield high quality and repeatability (Cary, 1995 and Santos, 1992). When the welding process is made in a multipass fashion (Lincoln, 1973) used for instance, in high thickness plates, the performance requirements for the automated systems are even tighter. The companies which can afford it, solve this problem with a robotic welding unit, constituted of one or more anthropomorphic robots with six degrees of freedom in general, indexing tables and the welding unit itself (Groover, 1986). Medium sizes companies, however, solve this problem with a lower cost solution composed of a displacement unit that drives the welding torch at a constant speed over the plates to be welded. This solution puts a heavy load on the operator who is required to monitor constantly the welding process for seam tracking and to adjust some of the welding parameters such as the height of the torch.

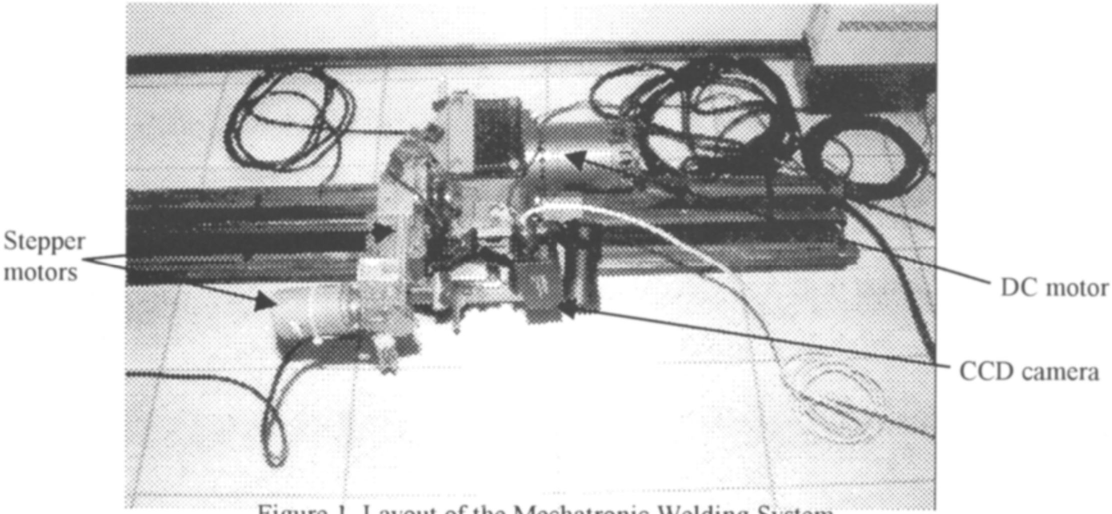


Figure 1. Layout of the Mechatronic Welding System

With regards to these two possible solutions, this paper presents an alternative, mechatronic solution for the multipass arc welding process which adds the essential monitoring features to the latter, with a cost that is far less than that of the former.

2. THE MECHATRONIC WELDING SYSTEM (MWS)

The description of the Mechatronic Welding System is based on the Figure 1: It is constituted of a mobile unit driven by a DC motor and two stepper motors attached to it, which are used for correcting the welding tracking in the horizontal and vertical directions The angular displacement of the motors is converted to linear displacement of the torch (vertical and horizontal) through a rack and pinion joint. The correction actions are triggered by a computer vision system whose CCD camera can be seen attached to the mobile unit. The vision system also inputs to the controller of the MWS the information related to the end of one welding pass and the starting point of the following one. Special mechanical and software filters are used with the optical components of the CCD camera in order to avoid light saturation in the CCD sensor.

A number of different aspects of the MWS could be described in further details. However, this paper focuses on the computer vision system, which is the major improvement that has been implemented in the MWS as compared to a commercial welding unit.

3. COMPUTER VISION SYSTEM

The main control actions taken by the MWS are triggered by the information obtained by a commercial computer vision system. The principal image processing algorithms are based upon border detection of the welding chord being formed, where the reference lines are obtained from: a 256 x 256 pixel grey image is initially binarized and then a pre-defined window is scanned for pixels of the line formed between the joint and the plate underneath it, as shown in figure 2.

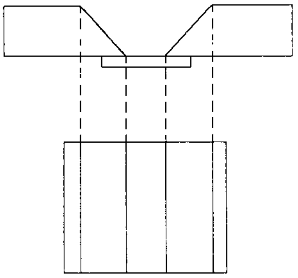


Figure 2: Line representation and detection

The upper part of this figure represents a lateral view of the plates to be welded and the bottom part shows the reference lines as viewed by the computer vision system, whose CCD camera is located over the plates joint, perpendicular to it. In order to avoid line discontinuity, a global analysis of the pixels is performed through the Hough transform (Fu et al, 1987). The lines obtained from these steps are used by the control unit to trigger several control actions which are described as follows.

3.1. Repositioning of the Welding Torch

The welding torch repositioning action is executed by two stepper motors in the horizontal (X) and vertical (Y) directions and two rack and pinion joints. These directions are related to repositioning of the welding torch at the beginning of a new chord (X) and welding stick out (Y). The choice of stepper motors for carrying out this task, was mainly based upon the fact that they work in a open loop control and their precision and accuracy were compatible with the actual welding requirements. The stepper motors are driven through a sequence of pulses which are sent to their control unit during a certain period of time (ΔT). The calculation of the variable ΔT is based upon the euclidean distance between the reference line of welding chord and the position of the welding torch that is, the distance in pixels obtained by the vision system is transformed into period of time over which the stepper motor should keep moving. To that, the following equation is used:

$$\Delta T = \frac{d}{k_1 . k_2 . k_3} \qquad (1)$$

where:

ΔT = period of time for driving the stepper motor;
d = distance in pixels between a given reference line and the actual position of the torch;
k_1 = angular displacement of the motor to time ratio [rad/s];
k_2 = pixels to distance (millimetres) ratio [pixels/mm];
k_3=linear to angular rack displacement ratio [mm/rad].

The constants k_1 to k_3 are obtained from off-line calibration procedures of the MWS system.

The sequence of repositioning steps of the welding torch is explained with the aid of figure 3(a to c).

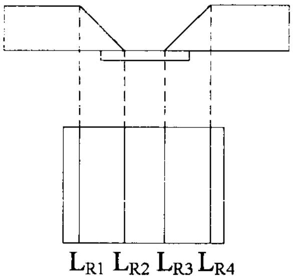


Figure 3.(a): Repositioning of the torch – Step1

Repositioning of the torch: step 1 (fig. 3.a)

- Detection of the reference line located at the left hand side of the image, which is the border of the chamfer (L_{R1});
- Detection of L_{R2} to L_{R4};
- Choice of L_{R2} to position the welding torch and
- Torch positioning

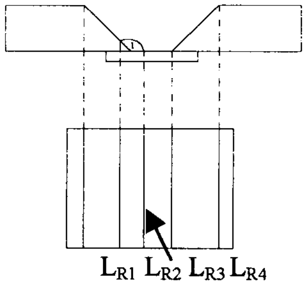


Figure 3.(b): Repositioning of the torch – Step2

Repositioning of the torch: step 2 (fig. 3.b)

- Welding chord #1 is made
- Mobile unit returns to its initial position
- Let L_{R2} (Step1) = L_{R1}(Step2)
- Detection of LR_2 and LR_3
- Choice of LR_2 to position the welding torch and
- Torch positioning

Repositioning of the torch: step 3 (fig. 3.c)

- Welding chord #2 is made
- The next steps are equal to those of the fig. 3.b, up to the point where the chamfered junction is completely filled with welding chords.

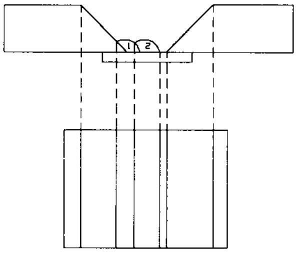


Figure 3.(c): Repositioning of the torch – Step3

3.2. Joint Tracking

The algorithm used in joint tracking is also based upon border detection, as it was for the torch repositioning; in this case, however, the borders are constantly detected by the vision system through successive image captures. By comparing the actual position of the welding torch with its future, known, position, a deviation measurement is generated. Figure 4 is used to explain this measurement: at a given time; the vision system finds the deviation (D) between the two reference lines. Up to point A (see figure 4), there are no deviations. After point B a deviation is found and the new reference line is taken to control the welding torch repositioning.

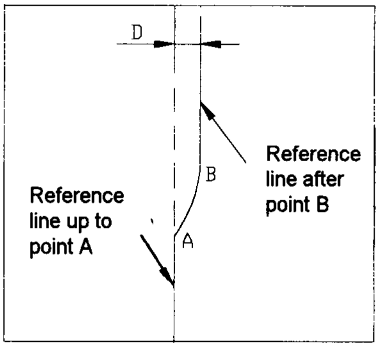


Figure 4. Joint tracking

However, when the detected deviation is lower than a limiar value(around .0.5 mm) this do not cause the control system to drive the motor to correct the welding tracking. The limiar value is set according to the quality requirements of the welding process. Similarly to the welding torch repositioning algorithm, this one also calculates the time interval during which the stepper motor should be operated.

3.3. Auxiliary Algorithms

The tracking and welding torch positioning algorithms are the main algorithms used by the control unit of the MWS. However, there are a number of auxiliary ones used to fully implement the control of the MWS. These are:

3.3.1.Tranversal Border Detection

This algorithm is used to find the stops of the welding process, which are located at the end and beginning of the plate which is being welded. It uses the same border detection algorithm used before. The only difference is the direction of the lines which it is seeking for: in this case, lines perpendicular to the joint lines (see figure 5).

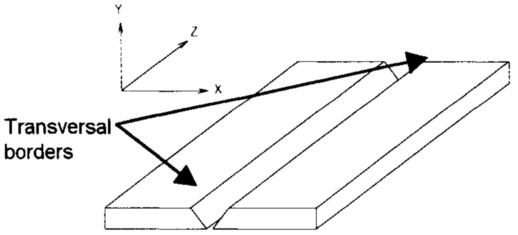


Figure 5: Transversal borders detection

3.3.2.Stick out adjustments

This algorithm is used for controlling the height of the welding torch, keeping the required stick out for a given welding process. The input for this algorithm is the position of the seam lines transformed by geometric considerations as shown in figure 6. The chamfer angle A is a known parameter of the welding process and the distance between the two reference lines $LR2_{[1]}$ and $LR2_{[2]}$ is calculated by the vision system so that the height of the welding torch is given equation (2).

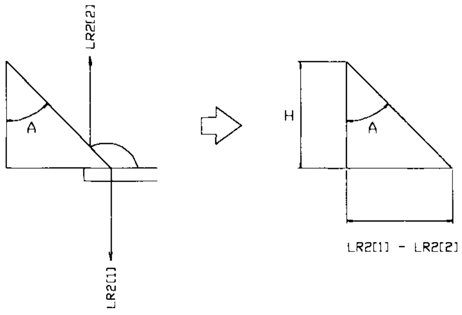


Figure 6. Stick out geometry

$$H = \frac{(LR2_{[1]} - LR2_{[2]})}{\tan A} \quad (2)$$

Once the height H is calculated, a period of time during which the stepper motor is activated to adjust the torch height, is calculated in a similar way of that presented in equation (1).

4. EXPERIMENTAL RESULTS

A number of experimental tests were conducted with the MWS unit in order to validate its operational characteristics. The main findings are reported below:

4.1. Repositioning of the welding torch (X direction)

After finishing a welding chord, the MWS is required to start a new one. See fig 3(b). A pen replaced the weldind torch in this test and the distance between its trace and the previous welding torch was measured a number of times and the average deviation was found. Its value was 0.35 [mm], which is an excellent result when one compares it with the average welding chord breadth that is 8 to 10[mm]

4.2. Repositioning of the welding torch (Y direction) –welding stick out

In order to carry out this test a number of specimen were prepared with different chamfer angles.

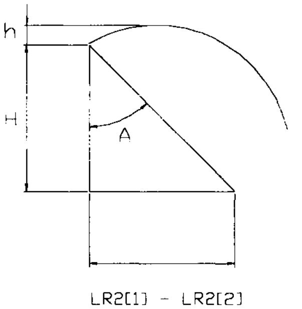


Figure 7. Geometry of the welding chord

It was found the heights calculated through equation (2) were always less than the required for the next welding chord. It has been concluded that a more precise way of calculating the height of the torch is given by the following equation:

$$H_T = H + h \quad (3)$$

where $\mathbf{H_T}$ is the correct value of the height of the torch, **H** is the height value obtained by equation (2) and **h** is the adjustment of the convexity of the welding chord as shown in figure 7.

After considering this new approach, where the value of the variable **h** was set by experimental tests, the height of the welding torch was found to be correct.

4.3. Joint Tracking

Usually, a deviation around 2 [mm] from the joint is tolerated for multipass welding applications. The tests conducted to assess this feature of the MWS unit an average error of 0.5 [mm], which shows a good performance of the unit with regards to joint tracking.

5. CONCLUSIONS

The Mechatronic Welding System which incorporates computer vision in its standard operation has got its major application in the arc welding process of large plates, which demand welding process in a multipass fashion. It is expected to have a big impact small and medium sized enterprises. The summary of its advantages takes into account two main areas: applicability and technical features.

Applicability:

- The low cost of the MWS fits very well the needs of medium and small sized enterprises;
- The operators action during the welding process practically does not exist. He (she) is required only to start the welding process up;
- The human welding skills are replaced by a systematic approach which is capable of making welding joints of steady quality.

Technical Features

- The use of computer vision to monitor and control the welding process makes it very flexible and reliable;
- The MWS was designed using the Mechatronic design paradigm that is the mechanical structure was designed right from the beginning to be controlled by a microcontroller.

Initial tests carried out with the Mechatronic Welding System yielded very good results: the multipass welding task was performed without any assistance of the operator. He (she) was only required to start the process up. The design and test phases of the MWS unit concluded so far point out to the complete success of this work.

REFERENCES

1. CARY, Howard B. Arc Welding Automation. New York: Marcel Dekker, 1995

2. FU, K.S., GONZALES, R.C., LEE, C.S.G., *Robotics: Control, Sensing, Vision and Intelligence. McGrall-Hill, 1987.*

3. GROOVER, M.P., WEII, M., NAGEL, R.N., ODREY, N.G., *Industrial Robotics: Technology, Programming and Applications,* . McGrall-Hill, 1987

4. Lincoln Electric Company, *The Procedure Handbook Of Arc Welding*. Cleveland 1973.

5. SANTOS, J.F.O., QUINTINO, L. *Automação e Robotização*. Edições Técnicas do Instituto de Soldadura e Qualidade, Portugal, 1992.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

The automated sorting of a bin of randomly placed documents of variable type and condition

C. Jones, G. Ross and J.R. Hewit

Department of Applied Physics and Electronic and Mechanical Engineering
University of Dundee, DD1 4HN, UK.

A random pile of paper items is sorted according to type and condition. To do this, it is first necessary to locate the topmost item so that it may be picked first from the pile. This requires the use of computer vision with algorithms based upon the Hough Transform. The edges of the items are identified using a specially arranged lighting system. Having located the position and orientation of the topmost item, a robot arm is programmed to pick the item from the pile and offer it to a series of condition monitoring devices. Finally, the item is placed on the appropriate one of a number of piles of similar items.

1. INTRODUCTION

Bin-picking by robot is a generic task common to a wide range of industries. Examples include the picking of piston rods from metal bins in a car engine plant, the picking of apples from boxes of fruit, and the picking of socks from a random pile at the output of a knitting machine.

As these three examples show, the bin-picking process can involve parts that are rigid and geometrically determinate, or parts that are flexible and deformable with little geometrical definition. The former are easier to bin-pick than the latter for two major reasons – their positions and orientations are easier to determine, and stay fixed after determination, and they are easier to pick up by a robot end-effector.

In all cases, however, it is likely that the same basic processes will be used to effect the bin-picking procedure. Computer vision will be used to locate the parts to be picked and, in particular, to differentiate between the part to be picked next and the other parts in the bin. A suitable end-effector will be used to do the actual picking. Tests may also be applied to the picked items.

2. PICKING OF PAPER ITEMS

There are many examples of industrial and commercial operations where the picking of paper items must be done accurately, robustly and fast. Banks must sort piles of bank notes of different denominations into other piles each containing only a single denomination; the Post Office must sort piles of letters and postcards prior to delivery; the counting of votes during an election is done by sorting the voting papers.

Traditionally, all of these tasks are done by humans who use visual and tactile sensing to locate the items and perform the classification. Where robots are proposed to replace the human it is essential that these sensory capabilities are replicated as far as technology will allow. For paper items, sensory and coordinative functions are used during the following repeated sequence -

- Find the topmost item on the pile
- Determine its position and orientation
- Use an end-effector to pick the item
- Apply tests on-the-fly to the item
- Determine whether it is defective
- Deposit the item in a suitable manner

These individual functions are described below.

3. MACHINE VISION ASPECTS

A random pile of paper items contained within a rectangular bin presents a bin-picking problem which is complicated by the variable form and type of the items. They may have any or all of the following: printed data on their surfaces, folded corners, torn edges or attachments such as staples or tape. The contents of the bin are viewed from above with a monochrome CCD camera and a typical pile as seen by the camera is shown in figure 3.1. Items comprising the pile are to be removed singly from the top of the pile. A wide range of items such as currency, for example, together with the possible inclusion of 'foreign' objects such as business cards, bus tickets etc. may be in the pile. This renders direct pattern recognition unfeasible and item location is derived from the physical boundary of items since the uppermost item will generally have its complete boundary visible.

3.1. Edge enhancement

Previously [1] thin and planar mechanical parts have been located in a bin-picking situation. These parts were located via edge tracking after grabbing a series of images illuminated from a range of positions. This application uses a similar lighting arrangement whereby we create shadows at the physical edges of the items by successively illuminating the pile from each side. A simple subtraction procedure would then, to a first approximation, remove the printed data. In reality the intensity of each image varies across each image. This variation should, in theory, be non-linear; however in practice it is found to be approximately linear. We therefore modify the images by taking on-line measurements of the intensity at each side of the bin, and apply a correction factor to each image. The final result after edge detection and binarisation is shown in Figure 3.2

3.2. Feature detection and location

The items constituting the pile are essentially rectangular but no edge orientation information is derived from the enhancement process and therefore the use of a generalised Hough transform [2] is not possible for object location. The fragmentation of physical edge features renders other methods such as boundary coding [3] unusable. We locate individual features of the items via a Hough transformation and have developed a simple accurate and robust procedure to locate these features. This is known as the **COA** algorithm[4]. The fragmentation of the edge data and its variable width together with the lack of edge orientation data produces a cluttered parameter space and non-ideal peak formations. The **COA** algorithm searches for regions indicating peaks in parameter space and is particularly effective when peaks are sharp [5] – as they are in this case. Having located these regions, the algorithm takes a slice through the peak at a level below the summit and computes the centre of area of this region. The effect of this is to ignore errors from distant parts of the distribution and also errors around the focal point of the peak. This can be likened to the Least Median of Squares approach to mode location [6]. We obtain sub-pixel accuracy of parameter estimation with little computational effort and the method has proved to be robust.

3.3. Object location

Final location of the topmost item of the pile is achieved by an analysis of the detected features based on a simple search of these features. Tolerances are incorporated on the angular location of each feature. The nature of the items within the bin and the preliminary image processing stages are such that very few items indicated in the binary edge image have truly parallel and perpendicular features. The complete image processing and analysis procedure incorporates many trade-offs, generally implemented via thresholding. Ultimately, the detection and location of the uppermost item is not guaranteed. However if the topmost item cannot be located, the contents of the bin are disturbed mechanically and the procedure is repeated until successful identification and location is achieved.

4. ITEM PICKING AND HANDLING

Individual items are to be picked from the bin and transported to sensing stations and stacked. A Staubli RX90 robot is utilised to provide the necessary six degrees of freedom. The automated handling of flexible materials [7] is a difficult task which in general requires a specially designed end-effector. The end-effector designed to pick and handle the wide range of items found in this application is shown in Figure 4.1

4.1. Picking

Attachment to the topmost item can only be made on its upper surface. Vacuum is used to make this attachment. The inner section of the end-effector houses two suction cups. The movement of the robot in the z-direction, when picking, is controlled by a vacuum sensor located in the line. The variable form of the pile renders optical methods of height detection unfeasible. Once attached to an item the end effector is lifted away and the inner section retracts. An ultrasonic transmitter is located within this extendible section for condition sensing.

4.2. Secure attachment.

The physical condition of the items is not known a-priori. Secure attachment is necessary so that accurate positioning over sensing stations can be achieved. It is also required because high dynamic forces may appear when transporting the items at high speed. High attachment forces are applied to each item irrespective of damage such as torn edges and porosity due to wear or damage, by application of vacuum to a porous pad. A high mass flow rate fan is used to supply the vacuum force which, with careful choice of the spacing and size of the holes, provides this secure attachment.

5. SENSING AND SORTING

Having located the top item of the unordered stack of items and picked that item, the document must be at least recognised if automatic sorting and placement are to be carried out. It may also be necessary, depending on the document, to validate it and screen it for condition. For example, a pile of currency notes might contain forgeries or notes too worn to be acceptable. The establishment of these properties would seem simple given that an imaging device is available within the system. This, however, is optimised for a very particular operation. To extend its uses to recognition etc. would require a separate illumination system, a different image set and further processing time. It is also unlikely that efficient validation and condition screening could be performed without using colour imaging. Recognition itself could prove time consuming for the existing system due to registration being only as good as the edge detection (± 2 mm laterally and ± 2 degrees skew). Thus, OCR would be difficult even for a small set of items.

The technique adopted here is to obtain data for recognition, validation and condition “on-the fly”. This leaves the imaging system to establish the next top document for picking when the end-effector returns.

5.1. Recognition

Recognition can be assisted by data on item size. The dimensions of items can be determined to within 2mm by the edge detect system. If the application is to sort several different types of items into classes and reject all others as a separate class, then it is possible first to reject all items whose dimensions are incorrect. However, even if each known class has unique dimensions this does not make its members unique in the universal document set. Some further form of differentiation is necessary. In this case the colour and spatial pattern along a fixed stripe are utilised to recognise known document types. Whilst in transit, the robot arm carries the item at a constant speed across a stationary colour sensor. This sensor has a wide bandwidth photodiode as the sensing element, with various LEDs positioned around the detector to provide known bandwidth illumination. The LEDs range from blue to near IR each with 40 nm bandwidth. The position of the LEDs and the aperture of the photodetector port are designed to exclude specular reflections from the item. The separate LEDs pulse sequentially and the reflections from the passing item provide a pulse train of colour information at a constant rate of 8000/m. Sample start/stop is automatically controlled and is dependent on the position of the end-effector and item dimensional information. The data is de-multiplexed into individual channel signatures which are compared using Bayesian classification (Eq 1) to a database from known items. Four stripes for each item are stored in the database to cover the four possible ways the item might be presented to the sensor. The closest fit is determined. Those items within tolerance are deemed to be recognised.

$$\boxed{\begin{array}{l} L_j = \sum_i [x_i - \mu_{ij}]^2 / \sigma_{ij}^2 \\ Class = j \rightarrow L_j = Min[L_j] \end{array}} \qquad (1.)$$

5.2. Validation

Depending on the tolerance of the recognition algorithm, most simple forgeries would be rejected at the recognition stage. This would include opportunist forgeries made by scanning originals and subsequently printing in CYMK on standard printer paper. The more valuable items such as currency, stamps, event tickets etc., are targeted by professional counterfeiters using lithographic printing techniques. To counteract these forgeries the system examines the signatures resulting from the colour differences and by applying time dependent discriminant analysis to the signatures. The discriminant functions are generated by training the system to highlight the locations of differences. This is shown in Figure 5.1.

5.3. Condition-screening

There are many aspects of an item covered by the term "condition-screening". It involves the item's aesthetic appearance, its age and the degree of wear, and the presence of folds tears and attachments. All of these features can be determined in various ways. Here we concentrate on age and degree of wear as expressed by the limpness of the item. Appearance is included implicitly, via the recognition and validation algorithms, but only for the specific path examined.

To determine limpness, ultrasonic transmittance is employed. A transmitter is located in the end-effector pad and at one point, during the passage of the item from the pile, the transmitter passes directly above a receiver in a known position. Processing of the limpness data must take into account the item recognition data. This is because limpness is subjective – for example, a limp business card may not be as limp as a new banknote

5.4. Sorting

Sorting is very application dependent. In this case rejects at any stage are sent to a specific reject bin for further mechanical or manual sorting. Items which are recognised and pass the condition screening process are sorted into ordered stacks within machine tolerance, or inserted into subsequent machines for binding or for redistribution.

6. RESULTS

6.1. Recognition

Using the Bayesian function (Equation 1) the system correctly identified 100% of the 5 types of item it was trained to recognise. The system also had a 0% false accept rate - i.e. all unknown documents were rejected

6.2. Validation

Tests for validation were carried out on genuine and copied items, where the copies ranged from colour scans and those from desk top colour printers to those that were printed using multicolour lithographic printing. In every case the copy was rejected as can be seen in Figure 6.1.

6.3. Screening

Condition screening was also almost entirely successful. It was found that the detected signal of a new item was substantially different to that of a used but subjectively acceptable one. Both were different again to that of a used and unacceptable item. These results have been obtained for all types of item investigated, though the signal levels were different for each of the document types examined.

6.4. Vision

The machine vision sub-system is three-stage (edge enhancement & detection, feature parameter estimation, feature analysis) with each stage itself being multi-stage. For example, the COA algorithm first locates peaks in parameter space and then estimates feature location. Many thresholds are required to be set to achieve successful object detection and location and each stage is highly dependent on the efficacy of previous stages. Sampling, using a wide range of items, resulted in success of detection and location as indicated in Figure 6.2. No complete failure to detect and locate was indicated, however, success was not necessarily achieved at the first attempt. Three modes of failure to detect and locate may occur:

1. There may be insufficient edge data for any particular edge feature to be detected.
2. There may be excessive noise in the image so that the search procedure registers occlusion.
3. The detected "edge" may be too wide so that the COA algorithm fails to locate any particular feature.

In the results obtained, where failures occurred, they occurred with the following frequencies -
Mode 1 - 10, Mode 2 – 4, Mode 3 - 7

REFERENCES

1. Dessimoz J.D., Kunt M., Zurcher J.M. Recognition and handling of overlapping industrial parts, Proc. 9th Int. Symp. on Industrial Robots, Washington DC, pp 357-366, 1979.
2. Ballard D.H., Generalising the Hough transform to detect arbitrary shapes, Pattern Recognition, 13, pp 111-122, 1981.
3. Freeman H. On the encoding of arbitrary geometric configurations, Trans. IEEE. Electron. Comput., 10, pp 260-268, 1961.
4. Jones C., Hewit J.R., The accurate and robust location of edge features of paper documents via the **COA** algorithm, (Submitted to Pattern Recognition Letters, 1998)
5. Davies E.R., Modelling peak shapes obtained by Hough transform, Proc. IEE, E, 139, 1, pp 9-12, 1992.
6. Davies E.R. Machine vision: Theory, Algorithms, Practicalities, Academic Press, 1997.

7. Taylor P. Handling of flexible materials in automation, Proc. Int. Workshop on Advanced Robotics and Intelligent Machines, University of Salford, UK, 1994.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Low-Cost Colour Vision-System for Robot Design Competitions

A.J.Baerveldt and B. Åstrand

Centre for Computer Systems Architecture, Halmstad University
Box 823 S-30118 Halmstad, Sweden

In this paper we present a low-cost colour vision system mainly intended for robot design competitions, which nowadays is a popular, project-oriented, way of teaching mechatronics in engineering curriculums. The estimated cost is about 450 dollar inclusive colour camera and the system is small enough to be carried on-board relative small mobile robots. The system is build around a signal processor TMS C31. We also present and discuss the experiences made with the system in our robot design competition.

1. INTRODUCTION

Most of the modern engineering products make use of some kind of "intelligence". This intelligence, often implemented on a microprocessor, is essentially what makes a product a mechatronic product. Today's engineers need at least a basic knowledge of mechatronic design techniques and elements such as sensors, actuators, mechanics, electronics, real-time software and computer hardware. However, to be able to really understand the interaction between the different elements and thus to obtain a *synergistic integration*, building a complete mechatronic system is a necessity. Therefore, we introduced a project oriented course in our master's degree program for Computer Systems Engineering with a specialisation in Mechatronics where the students can design and build a typical mechatronic product in the form of a mobile robot. The basic idea is that each group of two or three students gets a mobile robot kit to build and program a mobile robot with software to perform a predefined task. The task will be the same for all groups. At the end of the course the developed robots will compete with each other by performing the assigned task.

After running the course for three years, we felt that a major innovation of the course could be achieved if we could add a colour vision system to our robot kit. In this paper we first present the robot kit we use and discuss the experiences and shortcomings of our robot competitions. Then we present the low-cost colour vision system, which we specially developed for our course. Finally, we present the experiences of the first competition where our vision system was employed, which was in the spring of 1998.

2. ROBOT DESIGN COMPETITIONS

2.1 Motivation

The idea of a robot competition is not new and is for example since several years held at MIT in Boston and at ETH in Zurich. Having a robot competition in an academic curriculum is highly justified, as it confronts the students with open-ended problem spaces, forces to learn the students to

Figure 1 The robot kit, developed at MIT, is distributed to each student group.

work in groups (of 2 or 3 persons), and stimulates the creativity. Moreover, learners are often best at creating knowledge and ideas inside their own minds when at the same time they are engaged in an act of building a personally meaningful artefact in the world. It is our experience that the students really regard the design of a mobile robot focused on a competition as very stimulating and personally meaningful, especially while they follow the whole way from idea till a well-functioning system. These pedagogical aspects of a robot design competition are more thoroughly discussed by Fred Martin in his PhD Thesis [1].

2.2 Robot Kit

The robot kit we use is developed at Massachusetts Institute of Technology (MIT), in the USA. Low-cost has been the major reason for our choice as every group should has its own robot (10 robots in our case). The MIT kit costs about 600 dollar. The mobile robot kit consists of an onboard computer, several sensors, motors, batteries and LEGO-building parts. The motor package consists of four DC-motors and one servo motor. The sensor package consists of several contact switches, photodiodes, photoresistors, light emitting diodes and infrared senders and receivers. The LEGO-building parts, such as gears, wheels, rotational joints, chains, plates and beams, are provided to build the mechanical components of the robot. The computer board is built around a 68HC11 microcontroller and is equipped with 32 KB RAM together with the necessary I/O components to be able to drive the motors and read the sensor values, including H-bridges and AD-converters. The computer board can be programmed with a real-time operating system called Interactive-C, which is based on the programming language C. The kit is shown in figure 1.

As we found the sensor package rather limiting we added for example a distance sensor able to measure between 10 and 80 cm. We employed a SHARP-sensor based on triangulation available at surprisingly *low cost* and small size (about 25 dollar). This sensor proved to be very appropriate and due to its small field of view compared to ultra-sound very suitable to locate objects like golf balls.

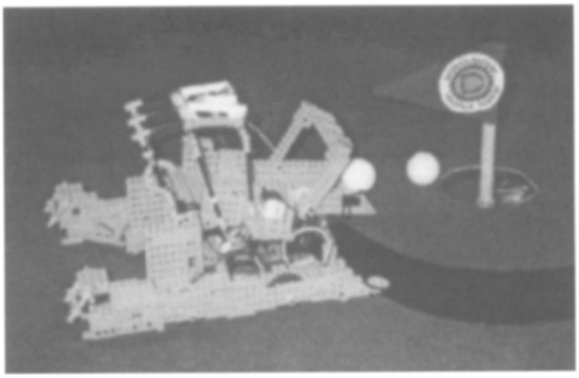


Figure 2 One of the robots delivering the golf balls at the green, trying to reach the hole.

2.3 Example Competition : Robot Golf Open

To give an example of a robot competition we will now describe the 1996 robot contest "Robot Golf Open". The contest arena was a rectangular square of 2 by 2 meter, surrounded by a 15 cm high wall. In the middle of the arena the so-called green was situated, which was a 7 cm high disk with a diameter of about 40 cm. The green contained the hole as can be seen in figure 2. Seven golf balls were randomly placed on the arena. It was the task of the robot to locate the balls, pick them up and put them somehow in the hole, which gave 2 points for each ball. 1 point could be gained if the ball was placed on the green only. It is emphasised here that the robots performed the task *autonomously*, i.e. based on the sensory information it made continuously decisions how to control itself, according to the software running on the on-board computer. Two robots "played golf" against each other for a period of two minutes.

The robot used the distance sensor to search for balls. The green could be found with a phototransistor, as a light was placed at the green. Contact switches informed the robot when it established contact with the green, collided with the surrounding wall or with the other robot. The competition was exciting to look at as all robots were able to place at least one ball in the green but often three or four.

2.4 Experiences and Shortcomings

At Halmstad University the first course started in 1995 and up to now we have held four competitions. Our experiences from the course is that the students are very attracted by the course, and especially the competition day has become a known event with around 250 spectators. The fact that the students are able to show their robots for so many people and the fact that the daily newspapers, local radio and TV used to report about the competition, strengthens their self-confidence and their motivation for the course.

The students really learn to integrate the different components of a mechatronic system and to apply and integrate their knowledge previously learned in the curriculum. They also experience the limitations of the sensors. Things that works in theory do not always work in real life. The use of LEGO also appeared to be very successful as it is easy to build and rebuild the mechanical parts of the system. The drawback we experienced with the MIT robot kit is that the sensors are quite limited to perform intelligent processing. Therefore, we soon added more advanced sensors, like the distance sensor mentioned before. Still the perception capabilities of the robot were quite limited and thus the possible tasks are limited in their complexity.

3. LOW-COST COLOUR VISION SYSTEM

3.1 Motivation

A computer vision system would be very appropriate to enhance the perception capabilities significantly. Especially, if a colour camera could be employed a lot of different objects can then be distinguished and located while using straight forward image processing algorithms. This would allow the design of very challenging robot competitions which require a fair amount of intelligent processing. Unfortunately, computer vision systems are in general relatively expensive, do not fit easily on a robot of about 30 by 17 cm and require a lot of computer power. However, due to the recent popularity of multimedia and Internet, the prices of small colour cameras with the necessary AD-conversion integrated, have now reached an acceptable level for our purpose. Connectix for example offers nowadays a colour camera with a parallel interface and with the size of nearly a Ping-Pong ball for about 250 dollar.

3.2 Overview

The vision sensor in our system is a colour camera from Connectix, called QuickCam Color II and is a very popular camera on the PC market due to low price and good image quality. To be able to run the computer vision algorithms sufficiently fast we decided to employ a separate processor for this purpose, namely the TMS C31 digital signal processor from Texas Instruments. This processor communicates with the camera and runs all computer vision algorithms. The final results from the image processing can then be transmitted to the 68HC11-board through a parallel interface. All decisions concerning the control of the actuators and the interpretation of the other sensor data are performed by the 68HC11-board. Thus the DSP-board together with the camera can be regarded as a *smart sensor.* This is illustrated in figure 3, where the system overview is given. The PC is only used during development.

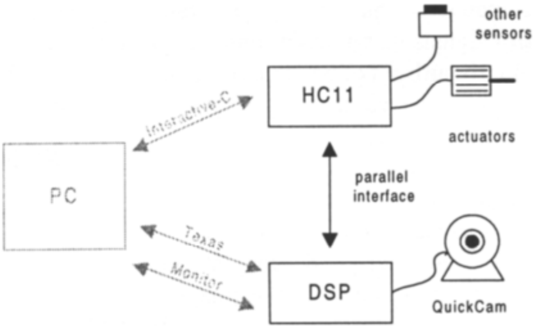


Figure 3 System overview with all major communication links. The parts and links in grey are only used during development.

3.3 Hardware

The QuickCam camera is equipped with a parallel port interface, thus a framegrabber becomes superfluous. This leads however to relatively low framerates due to the limitations of the interface. For example, it takes 0.3 seconds to transmit one RGB-image of 160x120 pixels. The images are in the well known RGB-colour format, with a maximum resolution of 320x240 pixels. A rectangular region of interest can freely be chosen anywhere in the image, as illustrated in figure 4. There is also a possibility to obtain sparse images, for example a decimation mode of 2 leaves out every second pixel thus leading to an image of 160x120 pixels [2]. So the problem of the limited communication bandwidth between the camera and the DSP can be overcome by reducing the amount of pixels to be transferred.

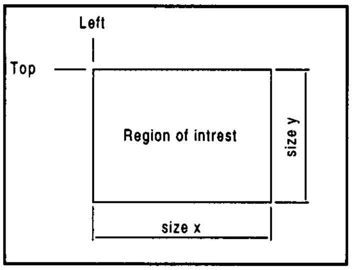


Figure 4 A region of interest can be chosen freely.

The computer system, belonging to the vision system, is built up around a single board computer equipped with a TMS320C31 32-bit floating-point DSP. Due to its very competitive price of 99 dollar, we use the so-called *starter kit* from Texas Instruments. It has pinned-out data, address and control signals that allow a simple expandability with daughterboards. To be able to connect the DSP-board with the camera, with the HC11-board and with the PC for development purposes, we designed our own daughterboard equipped with two additional parallel ports, eight binary inputs and eight binary outputs. We also put 64k word expansion memory on the daughterboard for storing the image processing programs and the image data. The total cost of the vision system is about 450$. 250$ for the camera, 99$ for the starter kit, and about 100$ for the daughterboard.

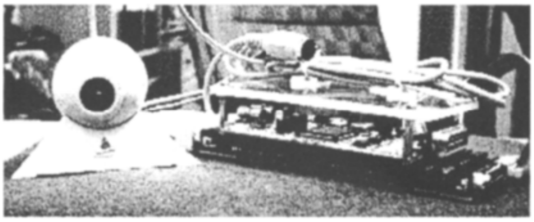


Figure 5 The low-cost colour vision-system with an estimated cost of 450 dollar.

3.4 Development Environment

The development of the image processing software by the students can be performed in two steps, whereof the first step can be omitted or performed with another software package.

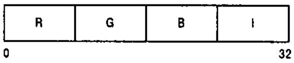
1. The RGB-images from the camera can be imported to MATLAB and analysed with the image processing toolbox [3] and with user written functions In this way, different algorithms can be tested efficiently.

2. Then the most suitable image processing algorithms can be programmed in C and downloaded to the DSP. The results can be displayed on the PC screen through a special monitor program.

Included in the starter kit, programs for assembling, downloading and debugging programs are provided. We have added a commercial C-compiler to be able to program the DSP in a high-level language. We wrote a special program which runs on the PC and communicates with DSP, to be able to control the camera and to display the images and the results on the screen of the PC. This program gives full access to all programmable functions of the camera.

3.5 Software

In this section we will describe which C-functions are available on the DSP. To start with, we have developed all interface routines for grabbing images at a desired resolution and region of interest as well as set and get all possible camera parameters such as brightness and contrast. Secondly, we provide routines to store the images in a buffer. One RGB-pixel is stored in each data word. When the image is grabbed the intensity for each pixel is calculated by the DSP and stored in the upper eight bits, as illustrated in figure 6. Finally, we also provide communication routines to communicate with the 68HC11-board.



$I = (R + G + B) / 3$

Figure 6 Contents of a 32-bit word in the image buffer.

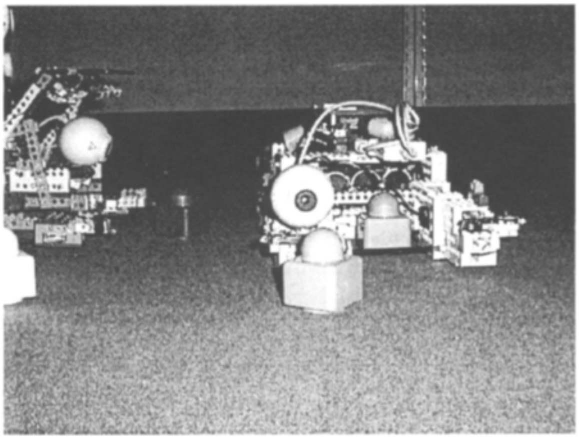


Figure 7 A LEGO-robot detecting a blue DUPLO-brick.

4. EXPERIENCES WITH THE VISION SYSTEM

4.1 The Robot Competition

The first contest with the vision system, held in spring 1998, was called election campaign and the task for the robots was to locate "voters" in an arena and bring them into the voting station. The arena was 2x2 meters and contained a ”voting station” in two of the corners, one for each robot. The voters are DUPLO-bricks and there are three different types of voters distinguished by the colours red, blue and yellow. The goal of the robot is to place as many voters, either red or blue, as possible in their own voting station, which is marked either with a red or a blue colour. This is illustrated in figure 7 and 8. Yellow voters give no points. Their roll is to act as tip of scale if the two teams collect the same number of voters.

4.2 Image Processing

The robot task required the robust segmentation of different colours. A well-known way [4] to achieve this is to normalise the RGB-colour signals with the intensity signal I, according to the following operations, which are performed for every pixel:

$$I = (R + G + B) / 3 \qquad (1)$$

$$r = R / I \qquad (2)$$

$$g = G / I \qquad (3)$$

$$b = B / I \qquad (4)$$

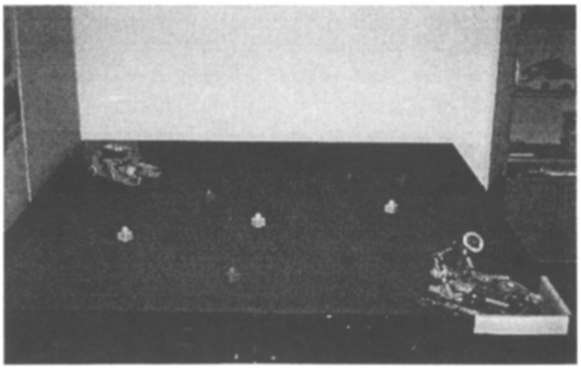


Figure 8 The arena, with the voting stations in the upper left and lower right corner.

Since each of the normalised components is linearly dependent of the other two, the normalised colour space can be represented by two normalised colours for example red (2) and blue (4). Each possible colour has its unique position in this two dimensional space, which is independent of the light intensity. This makes this method robust against variations in the intensity due to for example shadows as it only focuses on the real colour of the object. For the segmentation a minimum distance classifier can be used. Every RGB-pixel in the image is then converted to a normalised colour pair r,g and the distance is calculated to the predefined normalised colour of the object (r0,b0), according to:

$$d = \mathrm{SQRT}\,((r - r0)^2 + (b - b0)^2) \qquad (5)$$

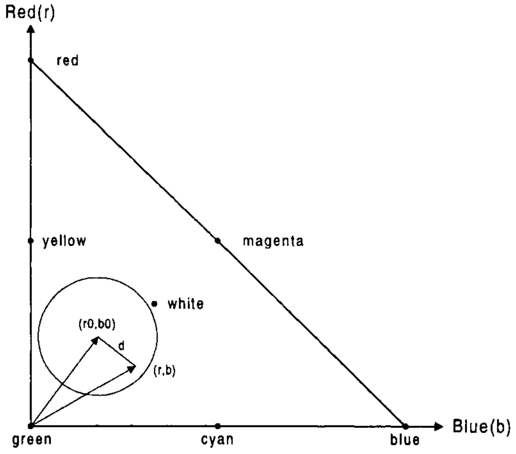


Figure 9 Normalised colour space.

If the distance d is smaller than a predefined threshold, the specific pixel is labelled as belonging to the object, otherwise it is a non-object point (see figure 9). The resulting binary image can be used for further processing, for example to calculate the area or the position of the area centre of the object. Dark pixels can give incorrect results, because the intensity signal I is almost zero and the normalised colour is mainly determined by noise. It is also possible that single bright pixels are misclassified due to specular reflections and other sources of noise. These misclassified pixels can be easily removed by some kind of filtering like a median filter.

4.3 Experiences

After intensive evaluation, the students employed either the above described method of colour segmentation or performed the minimum distance classifier directly based on the average RGB-value of the objects. Major differences occurred in the processing stage of the binary images. One group develop a function for *self-calibration* of the brightness to be able to adapt to different illuminations. Another group entered the area of *sensor-fusion* and combined information coming from the vision-system and distance-sensor effectively to detect and locate the DUPLO-bricks in the scene. Yet another group was able to reason about the distance of the DUPLO-bricks relative to the camera based on *perspective geometry*.

The students found the system very easy to work with. Students with no major experience of computer vision were capable to get very good results and learned a lot about the basic principles of computer vision. They also learned the difficulties of employing computer vision when the illumination cannot be controlled, which leads to for example frequent occurrence of specular reflectance. They were also extra motivated for the course as they experienced the integration of the vision system as being on the frontier of robotics.

The vision system proved to be compact enough to be carried onboard the relatively small robots. A typical execution time of 0.4 sec per image of 160x120 pixels was achieved, including grabbing and storing a frame (0.3 sec) and running the segmentation and localisation algorithms (0.1 sec). This allowed the robot to be controlled in real-time, i.e. the robot moved smoothly around while executing the vision algorithms.

5. CONCLUSIONS

Robot design competitions constitute nowadays a popular and effective way of teaching mechatronics in engineering curriculums at several universities. We designed a low-cost colour vision system mainly intended for such robot design competitions. The estimated cost is about 450 dollar inclusive colour camera and the system is small enough to be carried on-board mobile robots with a typical size of ca. 30 by 17 cm. The system is build around a signal processor TMS C31. The experiences with the system are very good and we are confident that the system will allow us the design of very challenging robot competitions in the future which require a significant amount of intelligent processing.

REFERENCES

[1] F. Martin, "Circuits to Control: Learning Engineering by Designing LEGO Robots", PhD Thesis, MIT Massachusetts Institute of Technology, June 1994.

[2] QuickCam User Guide, Connectix Corporation, USA 1996.

[3] Matlab, Image Processing Toolbox, The MathWorks Inc, USA May 1997.

[4] A. Gunzinger, S.Mathis and W. Guggenbühl, "Real Time Color Classification", 3rd Int. Workshop on Time-Varying Image Processing and Moving Object Recognition, Italy, May 1989.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

EXPLORING SPHERICAL IMAGES PROPERTIES FROM OFF-THE-SHELF CAMERAS

Inácio Fonseca, MSc Student
ISEC- Instituto Superior de Engenharia de Coimbra
Quinta da Nora, 3030 - COIMBRA, Portugal
e-mail: inacio@isec.pt, sousa@isr.uc.pt

Jorge Dias, PhD
I.S.R.-Institute of Systems and Robotics, Electrical Engineering Department,
University of Coimbra, Pinhal de Marrocos, 3030 - COIMBRA, Portugal
e-mail: jorge@isr.uc.pt

ABSTRACT

This paper addresses the implementation of a mechatronic device to explore spherical image properties using off-the-shelf cameras. The article describes the improvements and extensions performed on a first prototype built with mirrors to simulate patches in a spherical image surface. This first prototype was used as sensor on the navigation control module of a mobile robot. The navigational module controls the steering and forward movements of the mobile platform using visual information as feedback. The mirrors are adjusted to acquired images in different positions in the sphere. The success of the first prototype version inspired us to improve the system for a new version, where new image sphere properties could be better explored.

1 INTRODUCTION

This article presents ours efforts to develop a mechatronic device to explore spherical image properties, using off-the-shelf cameras. The developed prototype is described, as well as its features. Based on this prototype, a set of experimental results for mobile robot navigation are presented. During these experiments some interesting image sphere properties are explored.

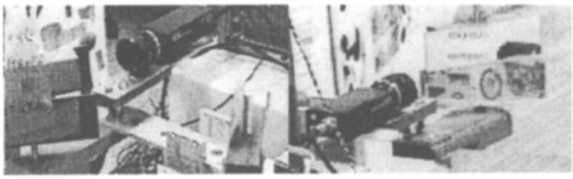


Figure 1. The actual system with the mirror setup to simulate four points of view in the sphere surface.

In the last section we present a new mechatronic device, that improves the initial prototype and enables the exploration of other image sphere properties.

Several solutions have been proposed by different authors to drive a mobile robot indoors (see [5,6 and 7] for some examples), but here we propose a different solution, based on image sphere properties. In the next section we introduce a model for the image sphere, analyzing it mathematically.

2 IMAGE SPHERE MODEL

Suppose that we want to control the attitude of a vehicle in the space, while the vehicle is moving with translacional velocity. Considering the referential system represented in the Fig. 2 and modeling the visual sensor as a spherical sensor with a radius r, we conclude that only two rotational velocities are necessary to control - (ω_x, ω_z).

To obtain these two velocities we can use images and from them we could compute the velocities induced in the image sphere by the vehicle's motion. This solution arises different problems. The first problem is how to determine the minimal number of cameras and their best position in the spherical surface to estimate the necessary information to control the vehicle. Bergholm [1] gives a relationship between the flow measured in the spherical equator perpendicular to the direction of translation the depth. This relationship is obtained by analyzing the self-projected flow (ego-motion) in the spherical surface, setting *r=1* by convenience and without losing generality.

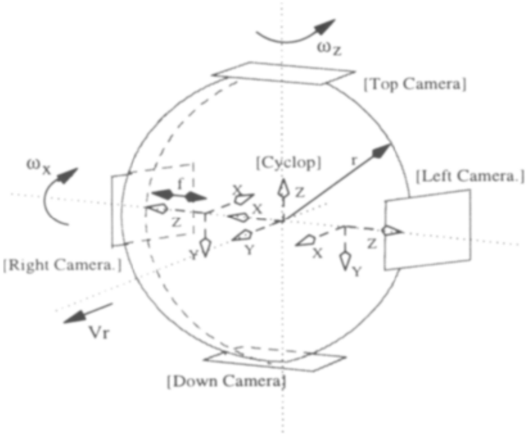


Figure 2:The sphere model for the navigation system in a three dimensional world.

The sphere's velocity can be described as $(\vec{v}, \vec{\omega})$ with $\vec{v}, \vec{\omega} \in R^3$. Let *d* be the distance from the sphere center to a three-dimensional point *X* with a vector $\vec{x}$ (the depth of the viewed point). Since $d = \sqrt{x^2 + y^2 + z^2}$, the following relation is valid:

$$\dot{\vec{x}} = \vec{\omega} \wedge \vec{x} + (\vec{v}/d) \cdot \|\vec{x}\| \tag{1}$$

It is also true, that any point *P* belonging to the *X* projection line can be described as $\vec{p} = \mu \cdot \vec{x}, \mu > 0$, and in particular if *P* belongs to the sphere surface then $\|\vec{p}\| = 1$. The *X* velocity projected in the sphere point P is then given by :

$$\dot{\vec{p}} = \vec{\omega} \wedge \vec{p} + (\vec{v}/d) \cdot \|\vec{p}\| \tag{2}$$

This velocity vector can be decomposed in any orthogonal referential, and in particular our choice will be the left referential represented in Fig. 3. This referential $[\vec{e}_x, \vec{e}_y, \vec{e}_z]$, enables the study of the image flow in the equator and locally. The point *P* described in this referential $[\vec{e}_x, \vec{e}_y, \vec{e}_z]$ has coordinates $[\cos(\theta), \sin(\theta), 0]$ where the vectors are given by $\vec{e}_y = [-\sin(\theta), \cos(\theta), 0]$, $\vec{e}_z = [0,0,1]$.

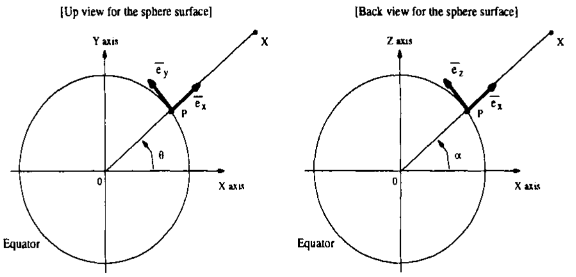


Figure 3: Decomposition of $\dot{\vec{p}}$ velocity in the referential described by the vectors $[\vec{e}_x, \vec{e}_y, \vec{e}_z]$.

Notice that the $\vec{e}_x$ is not important because we cannot measure this flow in the sphere surface. In this case we will have:

$$\begin{cases} \vec{e}_y \cdot \dot{\vec{p}} = \omega_z + \cos(\theta) \cdot v_y / d - \sin(\theta) \cdot v_x / d \\ \vec{e}_z \cdot \dot{\vec{p}} = \omega_x \cdot \sin(\theta) - \omega_y \cdot \cos(\theta) + v_z / d \end{cases} \tag{3}$$

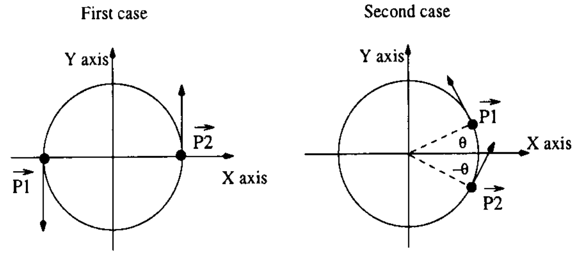
However, in our case the sphere velocities are given by $\vec{v} = [0, v_y,\ 0]$ (or $\vec{v} = [0, v_r, 0]$) and the rotational velocity is $\vec{\omega} = [\omega_x, 0, \omega_z]$. In this case the equation (3) reduces to:

$$\begin{cases} \vec{e}_y \cdot \dot{\vec{p}} = \omega_z + \cos(\theta) \cdot v_y / d \\ \vec{e}_z \cdot \dot{\vec{p}} = \omega_x \cdot \sin(\theta) \end{cases} \tag{4}$$

That is, with the equator normal flow $\vec{e}_z \cdot \dot{\vec{p}}$ we can estimate the ω_x velocity because it not depends from depth *d*. The equator parallel flow $\vec{e}_y \cdot \dot{\vec{p}}$ depends from ω_z but using the difference between the optical flow sensed in two points it is possible to remove it. Notice that the study described earlier, can be also done for the plane *y=0* and after doing the same analysis for the equator normal flow now $\vec{e}_y \cdot \dot{\vec{p}}$ and equator parallel flow now $\vec{e}_z \cdot \dot{\vec{p}}$ (now using the right referential in Fig. 3 by $[\vec{e}_x, \vec{e}_y, \vec{e}_z]$):

$$\begin{cases} \vec{e}_y \cdot \dot{\vec{p}} = \cos(\alpha) \cdot \omega_z - \sin(\alpha) \cdot \omega_x + v_y / d \\ \vec{e}_z \cdot \dot{\vec{p}} = 0 \end{cases} \tag{5}$$

From this analytic study we conclude that using the velocity measured in the sphere can be used to control the system. In fact comparing the optical flow from different points in the sphere is equivalent to compare the depth between two (or more) dimensional points in 3D space. This is the case when we compare divergent points in the sphere.


First case

Second case

Figure 4: Representation of the relevant points to control the ω_z velocity.

For simplicity let's first consider the ω_z control. From the Fig 4, and for the case of divergent points, the flow $\vec{e}_y \cdot \dot{\vec{p}}$ is,

- $\vec{e}_y \cdot \dot{\vec{p}}_2 + \vec{e}_y \cdot \dot{\vec{p}}_1 = 2\omega_z + v_y\left(\frac{1}{d_2} - \frac{1}{d_1}\right)$ which depends from ω_z. Subtracting we obtain:
- $\vec{e}_y \cdot \dot{\vec{p}}_2 - \vec{e}_y \cdot \dot{\vec{p}}_1 = v_y\left(\frac{1}{d_1} + \frac{1}{d_2}\right)$

From this equation we can not compute the depth between the two points. Looking now for the second case in the same figure and using two points of view from the same lateral position, where the flow $\vec{e}_y \cdot \dot{\vec{p}}$ is used:

- $\vec{e}_y \cdot \dot{\vec{p}}_2 + \vec{e}_y \cdot \dot{\vec{p}}_1 = 2\omega_z + 2\cos(\theta)v_y\left(\frac{1}{d_2} + \frac{1}{d_1}\right)$

which depends from ω_z

- $\vec{e}_y \cdot \dot{\vec{p}}_2 - \vec{e}_y \cdot \dot{\vec{p}}_1 = \cos(\theta)v_y\left(\frac{1}{d_2} - \frac{1}{d_1}\right)$

which enables to compare the depth between both points in presence of rotational movements ω_z and ω_x. From this analysis, we conclude that is possible to control the ω_z velocity if it is used four cameras: two for each lateral side of the image sphere. The control is also done in an independent way, i.e., we have a measure proportional to the orientation of each wall with the locomotion referential. For the control of ω_x we only need two cameras pointing as represented in Fig 2 by [Top Camera] and [Down Camera], because the ω_x value can be estimated by the flow given in the lateral cameras as explained earlier. So in this case using the equator normal flow $\vec{e}_y \cdot \dot{\vec{p}}$ from equation 5, the difference between the flow in the top and in the down side of the sphere is given by:

$$-2\omega_x + v_y\left(\frac{1}{d_1} - \frac{1}{d2}\right) \qquad (6).$$

The velocity ω_x could be removed by using the estimation given by the equator normal flow $\vec{e}_z \cdot \dot{\vec{p}}$ of the equation 4.

Notice that if we chose to measure the feedback signal only when the mobile robot is moving with linear velocity, the system only needs to have the four cameras and the best positions are those represented in the Fig. 2. With the left and right cameras the system can sense and control the ω_z velocity and with the others two the ω_x velocity. However with only four cameras it is impossible to control the system while doing angular movements, and to solve this problem it will be used two points of view for each lateral camera as explained earlier.

2.1 Cameras to Simulate a Spherical Sensor

In practice, is not necessary to have an image sphere to explore the properties described before. In fact we could use off-the-shelf cameras geometrically configured that acquire images corresponding to an image patch in the image sphere. To support the use of this principle we study this approach. Considering the setup in Fig 5, the $\vec{x}$ velocity in the sphere point $\vec{p}_s$ is $\dot{\vec{p}}_s = \frac{\dot{\vec{x}}}{\|\vec{x}\|}f$, but we can only measure in the sphere's surface the velocity $\dot{\vec{p}}_s = \frac{\dot{\vec{x}}}{\|\vec{x}\|}f - \frac{\dot{\vec{x}} \cdot \vec{x}}{\|\vec{x}\|^2} f \frac{\vec{x}}{\|\vec{x}\|}$.

For the camera, differentiating the geometrical projection law $\vec{p}_c = \frac{f}{\vec{x} \cdot \hat{k}}\vec{x}$ in order of t, we can express $\dot{\vec{p}}_c = \frac{f}{\vec{x} \cdot \hat{k}}\dot{\vec{x}} - \frac{f}{(\vec{x} \cdot \hat{k})^2}(\dot{\vec{x}} \cdot \hat{k})\vec{x}$, and using the

$\vec{x}$ velocity in this expression, we will get the final result that unfortunately is not equal to the measured velocity in the sphere's surface. However if the $\vec{x}$ vector points in the $\hat{k}$ direction the two velocities are actually the same.

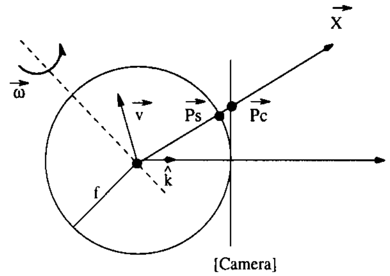
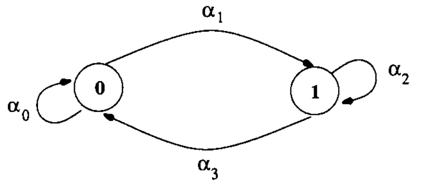

$\vec{x}$
$\vec{\omega}$
$\vec{v}$
$\vec{Ps}$
$\vec{Pc}$
$\hat{k}$
f
[Camera]

Figure 5: The conventional camera has the image plane tangent to a virtual spherical image sensor.

3 ALGORITHM AND RESULTS

3.1 Control Algorithm

In this section it will be described the experimental system used to test the approach proposed in last section. A controller based on the signals given by four images is designed to control the steering of a mobile platform. In this experimental system the mirrors are positioned in stereo divergent way, simulating the acquisition of images in the image sphere's equator.


α_1
α_0
0
1
α_2
α_3

$\Sigma_u = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$

Figure 6: The Discrete Event System definition. The states definitions are respectively: (State 0 – Acquisition, State 1 – Control). The event description is:

- α0 - wait until the next control timing arrives, i.e., 0,5s doing measures
- α1 - acquisition finished
- α2 - control update not finished yet
- α3 - command sent successful for the robot

To help the control and acquisition timings we use a DES (Discrete Event System) to monitor the algorithm presented in Fig 6. One of the states is responsible to monitor the acquisition and the other to monitor the communications with the mobile robot. Because the timings involved are different the final measure for each image is the mean in the acquisition interval. For steering control we use the lateral cameras, and two mirrors by camera. The signals, generated from optical flow obtained from the image, give a direct estimation the vehicle's attitude relative to the lateral planes. The signal measured in each image is the horizontal optical flow, which can be computed one-dimensionally using an optical-flow algorithm or a correlation technique. In our case the both are used to give a strong validation for the measure. The mirror setup is represented in Fig. 7.

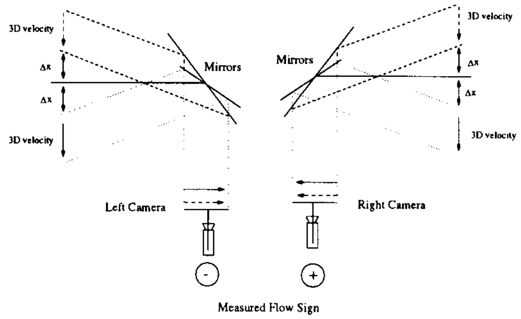

3D velocity
Δx
Δx
3D velocity
Mirrors
Mirrors
3D velocity
Δx
Δx
3D velocity
Left Camera
Right Camera
-
+
Measured Flow Sign

Figure 7: The mirror setup. In this case for simplicity the cameras do not have been represented as a geometrical model.

Considering the Fig. 7 and indicating the signals measured by:

- **FRMM** - Front Right Mirror Measure, optical flow measured in the front right image
- **BRMM** - Back Right Mirror Measure, optical flow measured in the back right image
- **FLMM** - Front Right Mirror Measure, optical flow measured in the front left image
- **BLMM** - Back Right Mirror Measure, optical flow measured in the back left image

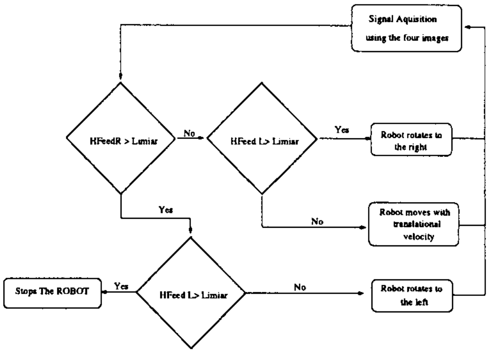

Signal Aquisition using the four images
HFeedR > Limiar
No
HFeed L> Limiar
Yes
Robot rotates to the right
Yes
No
Robot moves with translational velocity
Stops The ROBOT
Yes
HFeed L> Limiar
No
Robot rotates to the left

Figure 8: The control Algorithm diagram.

Defining *HfeedR* = *FRMM-BRMM* and *HfeedL* = *FLMM-BLMM* the algorithm can be described by the diagram represented in Fig. 8. In this case the forward velocity is constant and the system only controls the angular velocity. The timing for the acquisition loop is different from the control loop and they are respectively 0.08 sec and 0,5 sec.

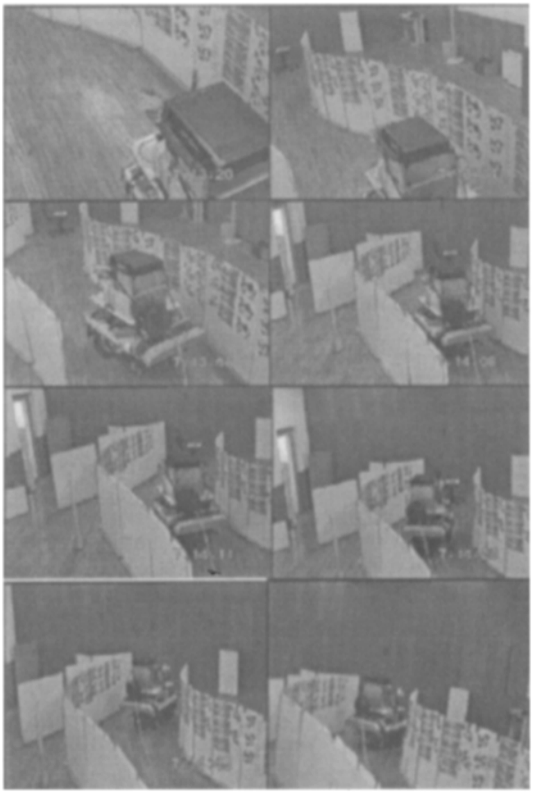


Figure 9: The result for the navigation with mirrors (external images).

The control strategy can be described as (just considering the right side):

- if $\|FRMM - BRMM\| <$ threshold then the robot can rotate to the left or do translations,
- if $FRMM - BRMM >$ threshold then the robot must rotate to the left,
- if $FRMM - BRMM <$ threshold then the robot can rotate to the left/right or going to the front (depends from the measures in the left side)

3.2 Experimental Results

The Figs. 9 shows the experimental results for the navigation process with mirrors and Fig 10 shows the measured signals to control the mobile platform.

4. SPHERICAL IMAGE SIMULATOR

The prototype that we plan to build in the future is represented in Fig 11,12 and 13. In this prototype, using a rotating mirror system and a rotating basis we simulate partially a small sphere's piece as showed in Fig. 12. In the construction we plan to use a camera pointing to the sky where is a mirror system similar to the one developed in the present work.

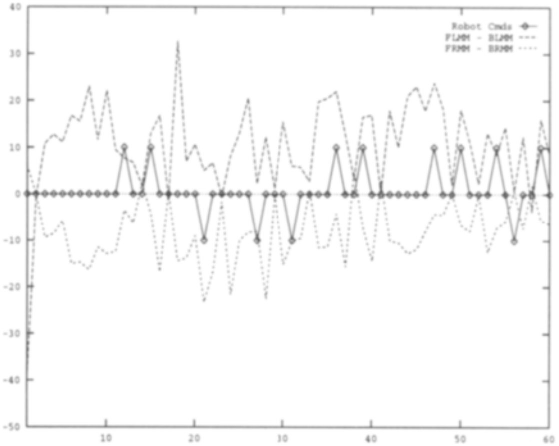


Figure 10:Feedback measures versus the commands sent to the mobile platform for the example showed in Fig 9. Time unit: 0,5 seconds. Commands: 0 - Forward movement, 10 - Right movement, 10 - Left movement.

The mirror system has two mirrors, one shows the left side and the other the right side. However each one can be moved by a motor that rotates it by using a two couple mechanical circles. The mirror system can be approximated to the camera by using the top motor. This is just do adjust the distance between the cameras and the mirrors (off-line procedure).

The basis can rotate and all the system rotates with it, including the camera. Comparing this prototype to the previous, we have few cameras, but a more complex mechanical system. At this moment we have not yet any idea about the size, we hope that the small piece that supports the mirror system can also support the two small motors that change the two mirror's orientation.

5. CONCLUSIONS

The paper addressed the problem of visual based navigation, using a mobile robot. The vision system realizes, partially, the concept of image sphere by using a stereo divergent system. To obtain more than two images in the sphere, mirrors were used. The experimental results obtained with the first prototype inspired us to continue with this approach of getting more accurate in the spherical sensor simulation.

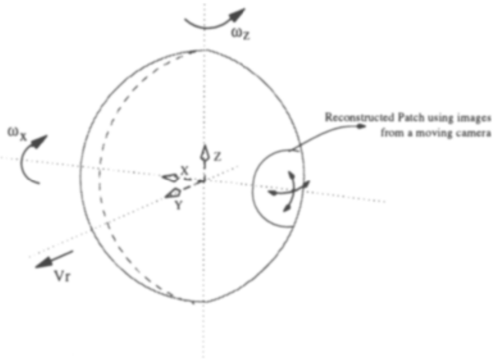

ωz
ωx
Z
X
Y
Vr
Reconstructed Patch using images from a moving camera

Figure 11: With the new prototype using a rotating basis and a rotating mirror system, we can cover the small circle represented.

REFERENCES

[1] F. Bergholm, "Decomposition Theory and Transformations of Visual Directions", ICCV 90-Int. Conf. on Computer Vision, Dec 1990, pp. 85-90.

[2] J. Dias, C. Paredes, I. Fonseca, H. Araújo,J. Batista, A.de Almeida, "Simulating pursuit with Machines - Experiments with Robots and Artificial Vision", ICRA'94, Int. Conference on Robotics and Automation,Nagoya, Japan, May 21-27, 1995.

[3] Jorge Dias, Carlos Paredes, Inácio Fonseca, Jorge Batista, Helder Araújo, A.de Almeida, " Simulating Pursuit with Machines, Using Active Vision and Mobile Robots", IEEE Transactions on Robotics and Automation, Vol 14, N° 1, February 1998.

[4] Roger, A. S. and Schwartz, E.L. "Design considerations for a Space-Variant Sensor with Complex Logarithmic Geometry", Proc. 10th International Conference on Pattern Recognition,Atlantic City, pp. 278-285,1990.

[5] Alberto Elfes, " Robot Navigation: Integrating Perception, Environmental Constraints and Task Execution Within a Probabilistic Framework", International Workshop, Ressoning with Uncertainty in Robotics Conf. - RUR, 95, Amsterdam, The Netherlands, December 1995, Springer ISBN 3-540-61376-5.Proc.

[6] Andrew J Davison and David W Murray, "Mobile Robot Localisation Using Active Vision", Parks Road, Oxford University, 1997.

[7] D. Coombs, M. Herman, T. Hong and M. Nashman, " Real-time Obstacle Avoidance Using Central Flow Divergence and Peripheral Flow ", National Institute of Standards and Technology, Nist Internal Report, June, 1995.

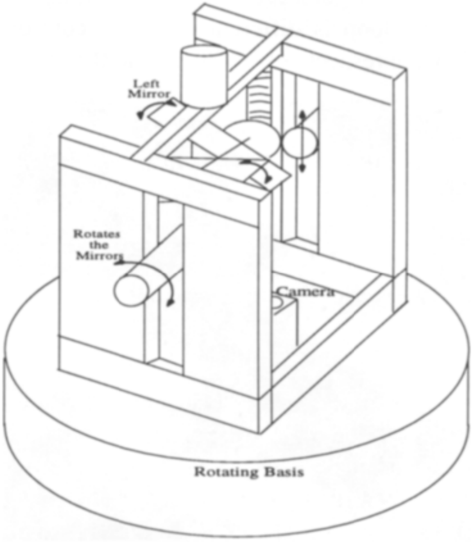

Left Mirror
Rotates the Mirror
Camera
Rotating Basis

Figure 12: The general view for the new spherical sensor.

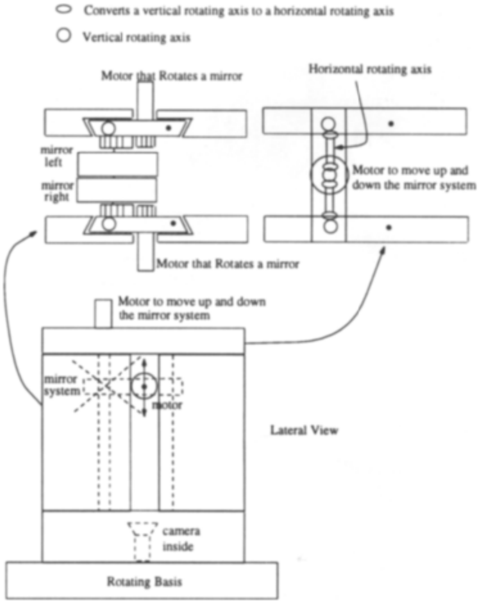

Converts a vertical rotating axis to a horizontal rotating axis
Vertical rotating axis
Motor that Rotates a mirror
Horizontal rotating axis
mirror left
mirror right
Motor to move up and down the mirror system
Motor that Rotates a mirror
Motor to move up and down the mirror system
mirror system
motor
Lateral View
camera inside
Rotating Basis

Figure 13: The prototype to the new spherical sensor simulation. This prototype uses a rotating basis (±90°) with only one camera, and a mirror system that can be rotated.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

COMBINING A FUZZY CLASSIFIER WITH CLASSIFIERS BASED ON STATISTIC MOMENTS*

Tito G. Amaral[a], Manuel M. Crisóstomo[b] and Aníbal T. de Almeida[b]

[a] Superior School of Technology - Polytechnic Institute of Setúbal,
Rua Vale de Chaves Estefanilha, 2910 Setúbal – PORTUGAL

[b] Electrical Engineering Department, University of Coimbra,
Researcher at the Institute of Systems and Robotics.
Pólo II, 3030 Coimbra, Portugal

This paper describes the behaviour of a global classifier constituted by a fuzzy classifier and two distance classifiers based on statistic moments. When we try to develop a classifier to recognise with successful a set of objects, we must have in mind that is impossible to obtain perfect performances (100% of recognition). However, performances between 80% and 95% are normally obtained with only one classifier. This fact, derive from the type and number of features used, which sometimes is insufficient to eliminate faults in the classification. It was developed a global classifier constituted by a combination of three classifiers based on two types of features. One is based on the visual perception such as area, perimeter and index of compactness, and the other two are based on statistic moments. The combination of the three classifiers is done using the voting principle.

The study was applied on binary images in a set of objects with various positions and orientations on the image plan. The study showsthat is possible, with good results, the combination of the fuzzy set theory with statistic theory on a multiple classifier.

Keywords
fuzzy classifier, statistic moments, global classifier, voting principle.

1. INTRODUCTION

Fuzzy logic was introduced by Professor Lofti Zadeh [1, 2] in 1965, as an extension to the traditional set theory. A fuzzy sets is determined by a membership function that maps members of a universe of discourse on to a membership range which is usually the interval [0, 1]. For a given element of the universe of discourse, if its membership value is 0, it indicates that the element does not belong to the set and if its membership value is 1 it indicates that the element represents completely the set. This theory allows a easier solution for some problems by modelling them according to the human reasoning and experience. In the recognition of objects, the problem arises from the efficiency and speed of recognition of objects. There is a trade-off between the number of features used for recognition and the speed of it. For better recognition it is recommended the use of a great number of features, but the greater this number is the slower the recognition process is. To

---

* Acknowledgements: this research was partially sponsored by JNICT through a scholarship (PRAXIS XXI/BM/1182/94) of the program PRAXIS XXI.

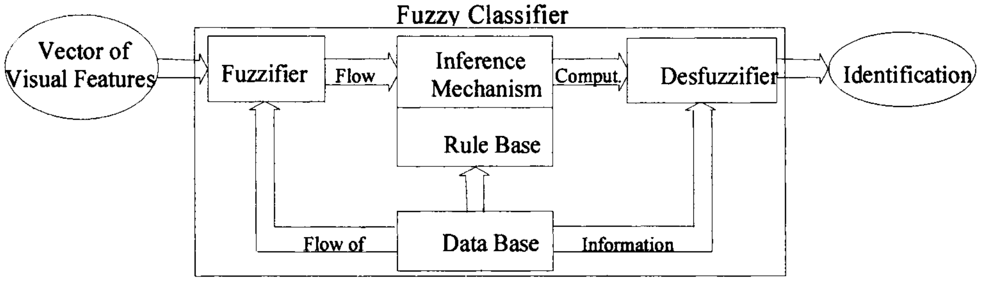


Figure 1. Fuzzy classifier.

overcome this problem it is important to define those features that better differentiate the objects and those that are redundant.

The traditional process of identifying objects is difficult due to factors such inherent noise and image distortion. These factors lead us to search for a reliable and robust approach in order to deal with ambiguous data of the images. An alternative seems to be the use of fuzzy logic classifiers and the possibility of combining with other type of classifiers, such as classifiers based on statistic moment invariants.

2. METHOD DESCRIPTION

It is presented the block diagram of the partial classifiers (a fuzzy classifier and two minimum distance classifiers) where the features used in the inputs are visual features and statistic moment invariants, respectively.

2.1 Fuzzy classifier

The visual features used in the fuzzy classifier are showed bellow [3]:

Area:

$$A(\mu) = \sum_{x=1}^{M} \sum_{y=1}^{N} \mu(x,y) \tag{1}$$

Perimeter:

$$P(\mu) = \sum_{x,y} \left(1 - \left|\mu(x) - \mu(y)\right|\right) * A_{xy} \tag{2}$$

Index of Compactness:

$$IOC(\mu) = \frac{A(\mu)}{P(\mu)^2} \tag{3}$$

Relation between the areas of the object and the area of its circular involving region:

$$\left(Circunf^2 \,/\, Area\right) = \frac{\left(\pi \times r^2\right)^2}{A(\mu)} \tag{4}$$

Distance between the centres of the object and of its circular involving region:

$$Offset(p_c, p_e) = \sqrt{\left[\left(\bar{x} - x_e\right)^2 + \left(\bar{y} - y_e\right)^2\right]}. \tag{5}$$

Other two visual features used are the number of corners of the object and his greatest side.

The block diagram of the fuzzy classifier is illustrated in the figure 1.

2.2 Distance classifier

We have used only the first four moment invariants in each distance classifier for numerical simplicity were the other moments are comparatively more complex and hence heavy in the computational loading. Another reason of using only the first four moments its because the higher moments are more sensitive to noise [4, 5].

The first four moment invariants of Hu are:

$$\phi 1 = \nu_{20} + \nu_{02} \tag{6}$$

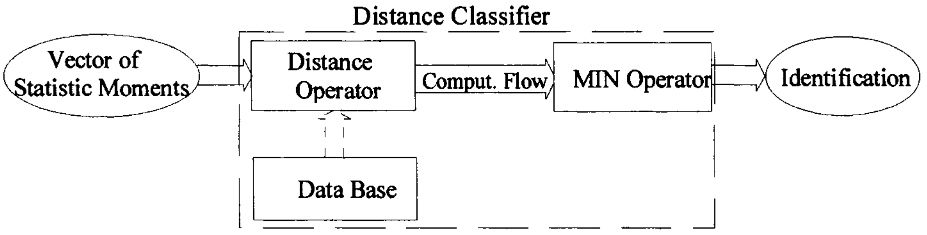


Figure 2. Distance classifier.

$$\phi 2 = \left(\nu_{20} - \nu_{02}\right)^2 + 4\nu_{11} \tag{7}$$

$$\phi 3 = \left(\nu_{30} - 3\nu_{12}\right)^2 + \left(3\nu_{21} - \nu_{03}\right)^2 \tag{8}$$

$$\phi 4 = \left(\nu_{30} + \nu_{12}\right)^2 + \left(\nu_{21} + \nu_{03}\right)^2 \tag{9}$$

The second distance classifier, is used on the first four moment invariants affine which are given by:

$$\psi 1 = \frac{1}{{\mu_{00}}^4}\left(\mu_{20}\mu_{02} - {\mu_{11}}^2\right) \tag{10}$$

$$\begin{aligned}\psi 2 = \frac{1}{{\mu_{00}}^{10}}\Big(&{\mu_{30}}^2{\mu_{03}}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + \\ &4\mu_{30}{\mu_{12}}^3 + 4\mu_{03}{\mu_{21}}^3 - 3{\mu_{21}}^2{\mu_{12}}^2\Big)\end{aligned} \tag{11}$$

$$\begin{aligned}\psi 3 = \frac{1}{{\mu_{00}}^7}\Big(&\mu_{20}\left(\mu_{21}\mu_{03} - {\mu_{12}}^2\right) - \\ &\mu_{11}\left(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}\right) + \mu_{02}\left(\mu_{30}\mu_{12} - {\mu_{21}}^2\right)\Big)\end{aligned} \tag{12}$$

$$\begin{aligned}\psi 4 = \frac{1}{{\mu_{00}}^{11}}\Big(&{\mu_{20}}^3{\mu_{03}}^2 - 6{\mu_{20}}^2\mu_{11}\mu_{12}\mu_{03} \\ &- 6{\mu_{20}}^2\mu_{02}\mu_{21}\mu_{03} + 9{\mu_{20}}^2\mu_{02}{\mu_{12}}^2 + 12\mu_{20}{\mu_{11}}^2\mu_{21}\mu_{03} \\ &+ 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\ &- 8{\mu_{11}}^3\mu_{30}\mu_{03} - 6\mu_{20}{\mu_{02}}^2\mu_{30}\mu_{12} + 9\mu_{20}{\mu_{02}}^2{\mu_{21}}^2 \\ &+ 12{\mu_{11}}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}{\mu_{02}}^2\mu_{30}\mu_{21} + {\mu_{02}}^3{\mu_{30}}^2\Big)\end{aligned}$$

where ν_{pq} and μ_{pq} are the normalised central moment and central moment, respectively, both of (p + q) order and are obtained through the following expressions:

$$\nu_{pq} = \frac{\mu_{pq}}{{\mu_{00}}^w}, \text{ with } w = \frac{p+q}{2} + 1 \tag{13}$$

$$\mu_{pq} = \sum_{i=1}^{M} I\left(x_i, y_i\right)\left(x_i - \bar{x}\right)^p\left(y_i - \bar{y}\right)^q . \tag{14}$$

The block diagram of the minimum distance classifier is illustrated in the figure 2.

2.3 Global classifier

The global classifier is obtained applying the voting principle in the outputs of the individual classifiers. Since the type of the outputs of the distance classifiers and fuzzy classifier it's not the same, then the combination of the classifiers is done in the order level. This level consists in the attribution of higher marks to the labels with minimum distance in the distance classifier and the greatest membership in the fuzzy classifier. The label which gets the higher marks is the output of the global classifier [6].

3. RESULTS

The input variables of each partial classifier are the extracted features in a set of objects illustrated in the figure 3. These objects are used for experimental work and it was assumed that they are not overlapped.

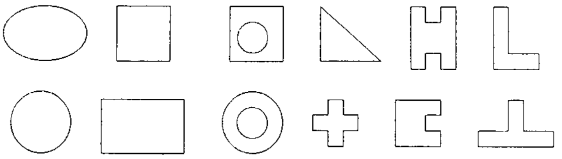


Figure 3. Objects used for identification.

Results show the future of the fuzzy sets theory in the classification stage and the possibility of combining this type of classifiers with those based on the statistic moment invariants. The block diagram of the global classifier is illustrated in the figure 4. Table 1 illustrates the recognition rate of each individual classifier and also the recognition rate of the global classifier. Table 2 ilustrates the processing time of each individual classifier.

Table 1. Recognition Rate of the classifiers.

Classifier	Recog. rate
Fuzzy (visual features)	95%
Distance (moment Hu)	83%
Distance (mom. affine)	71%
Global	94%

Table 2. Processing Time of the classifiers.

Classifier	Proc. Time (ms)
Fuzzy (visual features)	4,4
Distance (moment Hu)	12,1
Distance (mom. affine)	12,1

4. CONCLUSION

This paper describes a fuzzy classifier for object recognition and a global classifier based on distance and fuzzy classifiers. Experimental results show that identification of objects can be done with efficiency using the fuzzy classifier. The recognition rate of the distance classifiers was not so good because the moment vectors of the prototypes are too close and that causes incorrect classifications if the image contains some noise. The noise inherent to digital images was removed by means of a max-min filter. In consequence, high spatial frequencies were attenuated, resulting a smooth image.

The global classifier presents a recognition rate lower than the fuzzy classifier because the negative contribution of the based moment classifiers. Although his performance is very acceptable and demonstrate the possibility of combining classifiers of different types. The area of the object, its index of compactness, the relation between the area of the object and the area of its circular involving region, the distance between the gravity centres of the object and its rectangular involving region, the number of corners of the object and his greatest side are the features used in the input of the fuzzy classifier. In the distance classifiers was used only the first four moment invariants because the higher moments are more sensitive to noise. The use of other features in the fuzzy classifier such as aspect ratio, marginal distribution, and density of the involved area, should improve the identification process, but the computational cost should be paid.

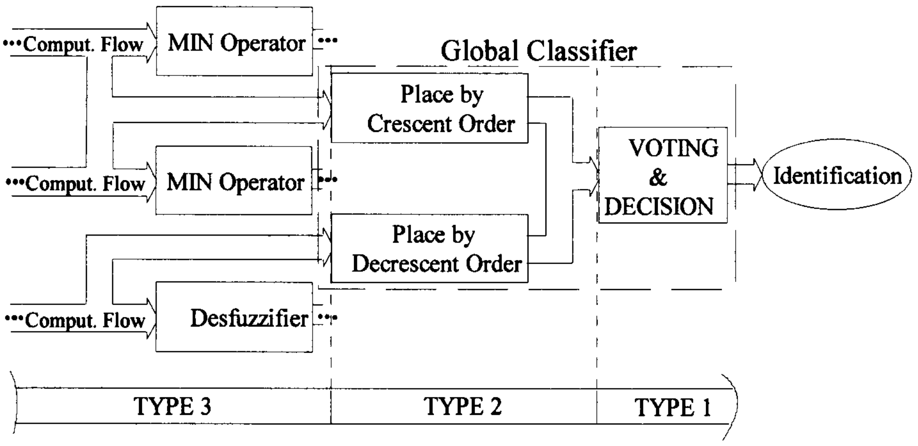


Figure 4. Global classifier.

REFERENCES

1. L. A. Zadeh, "Fuzzy sets", Inform. & Contr. Vol. 8, (1965), 338-353.
2. L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes", IEEE Trans. on Systems, Man & Cybernetics, Vol. SMC-3, (1973), 28-44.
3. Tito G. Amaral, Manuel M. Crisóstomo e Aníbal T. de Almeida, "Lógica Fuzzy aplicada ao reconhecimento de objectos em movimento", 2° Encontro Nacional do Colégio de Engenharia Electrotécnica, (Dezembro 1995), 145-152.
4. Jan Flusser and Tomás Suk, "Affine moment invariants: a new tool for character recognition", Pattern Recognition Letters, Vol. 15, (April, 1994), 433-436.
5. Jan Flusser and Tomás Suk, "Pattern recognition by affine moment invariants", Pattern Recognition, Vol. 26, (1993), 167-174.
6. Lei Xu and Adam Krzyzak, "Methods of combining multiple classifiers and their applications to handwriting recognition", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 22, N° 3, May/June, 1992, pp. 418-435.

.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Surface Profile Based on Sensor Fusion

Cristina Santos, Jaime Fonseca, Paulo Garrido, Carlos Couto
Department of Industrial Electronics
School of Engineering - University of Minho
Braga-Portugal

This paper reports a sensor system which it has been designed and constructed to acquire the profile of surfaces. This system is based in a CCD camera for object boundary-determination mounted over a robot manipulator shoulder and ultrasonic sensors for depth measurement mounted on a fixture at the wrist of the manipulator. It has been used an average weighted by degrees of confidence for raw sensor data fusion, based on a heuristic set of rules.

1. INTRODUCTION

From an economical point of view may be interesting to replace a single highly accurate but expensive sensor by several less precise low-cost sensors requiring with additional post-processing electronics. Using several low-cost sensors combined with intelligent post-processing can compensate the its low accuracy. These sensors can be either of the same type or give complement information. With the same type of sensors the goal is to increase the quality of the resulting sensor information. Of course, the improvement must be reasonable compared with the increasing complexity of the measurement system in order to keep the overall cost still attractive. As the computing power cost is decreasing everyday and the low cost sensors is bound to proliferate in the near future, multisensor systems and sensor fusion techniques bound to became more and more popular.
Several sensor fusion methods have been reported that deal with this problem. Durrant-Whyte has developed a Bayesian estimation technique for combining touch and stereo sensing[8]. Tang and Lee proposed a generic framework that employs a sensor-independent, feature-based relational model to represent information acquired by various sensors[19]. In[9], a Kalman filter update equation was developed to obtain the correspondence of a line segment to a model, and this correspondence was then used to correct position estimation. In[18], a extended Kalman filter was used to manipulate image and spatial uncertainties.

The work described in this paper is concerned with the development of a robot based system to acquire the profile of a surface. A PUMA 560 manipulator was equipped with a CCD video camera and one ultrasonic sensor in the shoulder and a set of four ultrasonic sensors in the wrist, to acquire data for internally representing the geometry of the part's surface, exploiting the mobility of the robot. The camera defines the work area while the sensors enable to get the surface profile. We used an average weighted by degrees of confidence for raw sensor data fusion to obtain more reliable representation in profile surface acquisition.

This paper is organised as follows: Section 2 describes our hardware configuration, Section 3 describes the software procedures which implement image acquisition and depth map filling of the part to be acquired, Section 4 presents tests and results and, Section 5 presents the conclusions and future work suggestions.

2. HARDWARE CONFIGURATION

2.1. Sensor System Design

The sensor system consists of a National Electronics monochrome CCD video camera and five ultrasonic sensors. The CCD camera and one ultrasonic sensor are mounted over the shoulder of the Puma, both horizontally aligned, to enabling a circular scanning around the robot stand. The other four ultrasonic sensors are mounted on a fixture at the wrist of the manipulator, hence benefiting from all degrees of freedom of the robot. The disposition of these ultrasonic sensors, relative to the robot grip's axis, is a square as presented in the figure 1. The distance between them is few centimetres.

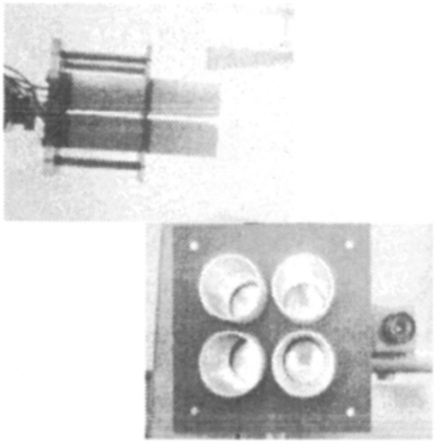


Fig.1 - Location of the sensors. a) Side view. b) Front view

The sensors used in this work are adopted from Polaroid Ultrasonic Ranging Units, which have a range of about 0.35m to 4m. The ultrasonic transducers have emission frequencies of 63kHz and are controlled by a specific kit provided by Polaroid Corp. This kit is based in the Intel 80C196 microprocessor and is easy to configure by software. It is possible to configure the following parameters: transmit frequency, pulse width, blanking time, amplifier gain, sample rate and trigger source internal/external. This kit is connected with external world by RS-232. To avoid the possible interference from the emission and echo waves, the sensors are triggered sequentially, leaving just one unit emitting at a time.

The CCD camera, with a resolution 500(H) x 582(V) is connected to the low-cost Creative's Video Blaster acquisition board inside of the Pentium PC. The software for the image processing was developed using Borland C under Windows environment.

2.2. Limitations of Ultrasonic Sensors

The basic principle of distance measurement using ultrasonic sensor is to record the time between the emission of the transmission wave and the reception of the returned echo. However, due to the wide beam opening angle of the ultrasonic transducers, a measurement error will occur if the emitted wave is not perpendicular to the target object. As depicted in the figure 2., the distance between the sensor and a flat wall should be $\overline{OA}$. But we will detect the distance $\overline{OA'}$, because this point provides the first returned echo. The position measured in this case is then given by A.

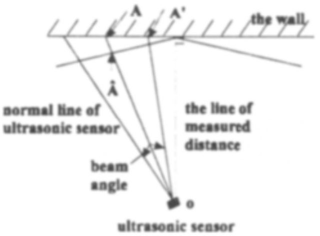

A
A'
the wall
A
normal line of ultrasonic sensor
the line of measured distance
beam angle
o
ultrasonic sensor

Fig. 2 - The influence of the beam opening angle when scanning a wall with an ultrasonic sensor

Therefore, there is a difference between the location of the perceived wall and the location of the actual wall. This error caused by the beam opening angle affects the accuracy of the environment perception. We found that the maximum beam opening angle for the sensors is about 22°. On the other hand, the optimal angle between the line of emission and the surface of the measurement object is 90°. However, if the incident angle exceeds the maximum value α, the echo may be received after multiple reflections. The detected data will then be erroneous. Therefore, ultrasonic sensors can easily measure the distance to an object, but they have poor angular resolution, due to the beam opening angle and to the spurious reflections phenomenon.

The methodology followed to decrease the beam opening angle was to increase the frequency (63kHz) and to put a tube 20cm long with 4cm diameter in front of the ultrasonic sensors.

2.2. Can Network

The computing hardware includes two CAN boards, the *Universal CAN I/O board* outside the computer and the *PC-CAN Interface PCI02* inside the PC. Both boards are based on the Intel 80592, products of STZP (Steinbeis Transferzentrum Prozessautomatisierung).

The *Universal CAN I/O board* deals with the Polaroid 's kit receiving the data sent and assuring the sequential triggering of the transducers. In reply to a trigger signal several measurements are made and the average value is calculated. This pre-processed data is then sent to the PC via the CAN net at 1Mb baudrate. This CAN board has the following features: 16 digital inputs, 16 digital output, 8 analogue inputs and 2 pulse with modulated output.

The software was developed in *IAR C* for the *Universal CAN I/O board* and in *Borland C* for the *PCI02 board.*

The software for communication is developed in IAR C and Borland C for the Universal CAN I/O board and PCI02 board.

2.3. PUMA 560

The PUMA 560 is a six-joint industrial robot manipulator, whose original UNIMATE MARK II controller includes a DEC LSI/11 processor and six digital and analogue servo boards. Interaction with the user is possible through the VAL-II operating system. The LSI/11 interprets VAL-II statements and generates joint interpolated co-ordinates, sending them to the six 6503 microprocessors. The PUMA 560 original control architecture is depicted in figure 3.

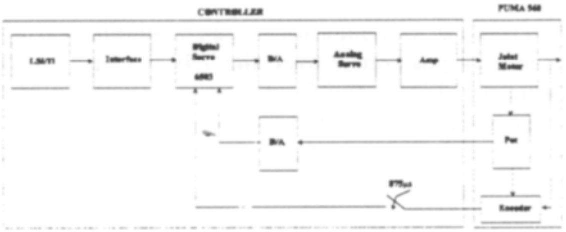


Fig. 3 - Original control architecture

This architecture offers many difficulties when used for high level control tasks, since the operating system VAL-II is based on a closed architecture. Resources to accomplish sensing, such as vision and force, are not supported. Also, the joint controllers have fixed gains and it is impossible to change them or to change the way the controller generates the paths following its internal interpolators. Hence, the implementation of new strategies for each of the joint controllers, the generation of trajectories or task planning algorithms are not possible.

The solution adopted to avoid the above mentioned limitations, was to replace the operating system VAL-II and most of the manipulator control hardware. The new installed hardware gives direct access to the joint positions and bypasses the old joint controllers. This new system is based on the *Trident Robotics cards* TRC004, TRC100, and TRC041 (*Puma cable card set*). The system composed by the TRC004 and TRC100 cards plus a personal computer (PC), replaces the old system composed by the LSI/11 processor, the VAL-II operating system and the joint controllers. The remaining components from the original architecture are the incremental encoders, potentiometers, power amplifiers and analogue servos.

The TRC004 is a general purpose interface board for servo applications. It provides eight channels of buffered analogue output, eight channels of analogue input, six quadrature shaft encoder channels, six bits of discrete input, six bits of discrete output and a watchdog timer. Physically, the CPU, RAM, EPROM, serial controller, interface cards, and joint servos are all removed from controller backplane. In their place, the TRC004, a single twelve-inch by nine-inch (approximately) multi-purpose I/O card is mounted and wired to the backplane by TRC041 card.

The TRC100 is a general purpose RISC processor board for servo control and data acquisition applications. With its local I/O bus the processor operates either standalone or in an ISA or EISA backplane. For real-time control, the code is either downloaded from a host computer or stored in EPROM. In our case, this card provides an interface between the TRC004 and the ISA bus of the PC.

The block diagram of the new architecture is represented in figure 4.

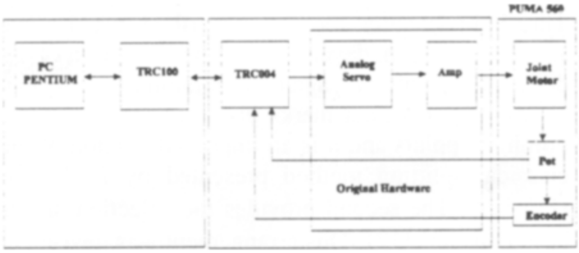


Fig. 4 - New architecture

The interface between the TRC100 and TRC004 is accomplished by a software developed by Trident Robotics (a DLL for Windows environment and some software placed in the EPROM of TRC100).

3. PROFILE SURFACE

In this section we will describe all steps needed to acquire surface profile: object search and robot positioning, surface scanning for depth acquisition.

The robot is positioned at the centre of a ring table, in which objects whose surface has to be acquired can be positioned. This table has 100cm of height, 95 cm of internal ratio and 125 cm of external ratio.

3.1. Search for the object and robot positioning

The incremental rotational movement of the robot's base and the processing of the acquired images allow the location of the object performing the search process.

After the object detection, the system stops the rotational movement of the robot and centres the object in the vision plan of the camera, as shown in the figure 5. Next, the dominant points of the contour are extracted in order to create a 2D representation of the part's surface.

The extraction of the dominant points is implemented by the use of the combination of two algorithms. The first algorithm performs segmentation, which is achieved by Otsu global thresholding [2], selected on the basis of a comparative study covering Otsu, Maximum Entropy, Uniform Error and Minimum Error Threshold selection methods described in [11]. The second algorithm, developed for the extraction of the dominant points, is again a combination of two algorithms. The first marks pixels as candidates for dominant points and it is an improved version of the classical splitting method presented by Duda and Hart [13]. The second provides the selection and is based on slope[17]. This arrangement was devised to provide a process for dominant points extraction suitable for most different sorts of object shapes. The dominant points are depicted in the figure 5.

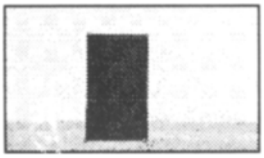


Fig. 5 – Object extraction from background: Dominants points

The method implemented for calibration allow the object to present the correct dimensions since positioned over the worktable. The ultrasonic sensor mounted over the camera measures the distance between the camera and the object's centroid.

The robot is moved to a fixed distance of the table at the table height and in the direction of the centroid's line.

The process described confines the work area of the manipulator, and sets the system ready for horizontal object scanning.

3.2. Surface scanning for depth acquisition

The 3D acquisition is accomplished by making the manipulator scan the 2D shape with its ultrasonic sensors. The overall result of this task is the building of a surface map that shall support the generation of profile surface. The y co-ordinates that the image acquisition obtained for the part's dominant points are subjected to perspective errors. Those values were based on the depth measurement given by the ultrasonic sensor mounted on the top of the video camera, which isn't the depth of all points but approximately of the object's centroid, as the figure 6.a) illustrates. The only co-ordinate that the system can consider without error is the z co-ordinate as the object is positioned over the table, where z co-ordinate is known and fixed.

This way, when acquiring points in the boundary region an uncertainty margin must be considered and the found boundaries are augmented by a certain default value, as figure 6.b) illustrates.

This contour enlargement minimises the perspective and calibration errors because the true values of x, y co-ordinates of the boundary points will be validated later by the scanning with ultrasonic sensors.

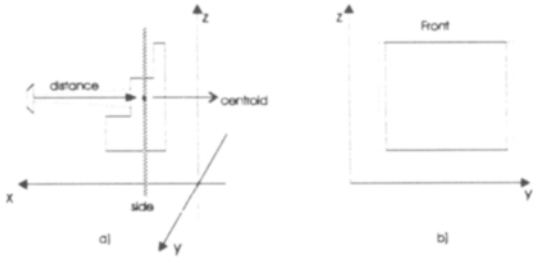


Fig. 6 - Object acquisition a) Side view b) Front view and contour enlargement

The algorithm implemented calculates the next position for acquisition using a fixed step. This step has the same value for the z and y co-ordinates. For each horizontal scan line, the start point is always defined by one extreme of the boundary calculated and the robot will step along evenly spaced points, till the end of the scan line. The definition of this step is done "a priori" and it depends on the desired precision for acquisition and the minimum resolution allowed to the surface. We used a fixed step s given by $s = \frac{ds}{5}$, where ds is the distance between the sensors.

In order to calculate the joint angles needed to position the robot in the next position, we used inverse kinematics since the manipulator's position is known. This feature is given by the new architecture that allows the necessary feedback to control the robot. For the implementation of inverse kinematics we used Borland C, based on the Robotics Toolbox of Peter Corke [14].

Assuming that the surface does not present significant changes in the square defined by the four ultrasonic sensors, it seems reasonable to accept that the average of the sensors measurements can be assigned to that square centre point of the part's surface. This point lies on the grip axis, i.e., in the axis passing through the centre of the sensor arrangement. Following this procedure, to maintain the resolution of the depth map, the speed of acquisition must be traded-off against sensor fusion in order to get greater accuracy. It is always a start-stop process, as the system needs the sensor data to decide what to do next.

Besides spatially averaging the ultrasonic readings, weights were introduced affecting each of the sensor measurements. These weights are to be interpreted as degrees of confidence, based on a heuristic set of rules [3] [5]. The sensor fusion problem is treated like if a group of human beings is positioned in front of a decision problem. For each point, after reading the data and the parameters defining the situation of acquisition, the system introduces a degree of confidence to each measure collected, where a high value denotes high confidence that the read value is correct. This degree of confidence takes values between [0,1] and can be interpreted as a probability given to the veracity of the proposition "The measurement read by the sensor i is correct".

The distance assigned to the surface's point lying at the grip axis for the scanning position n, was denoted M and is given by:

$$M(n) = \frac{\sum_{i=1}^{4} m_{sensor}(i) \times dc(i)}{\sum_{i=1}^{4} dc(i)}, \qquad (1)$$

where i is the identifier of a sensor, $m_{sensor}(i)$ the measurement given by the sensor i and $dc(i)$ its degree of confidence.

It is worth to note that the m_{sensor} values are validated before this calculation according to the range of values physically possible. A set of tests is applied to each measurement to determine if it corresponds to a point inside or outside the object and to testify if the received value is consistent with the physical set-up. Sensors whose measurements are not validated do not enter into the average calculation. The parameter dc is established as the result of the product of 3 partial degrees of confidence

$$dc(i) = dc\_1(i) \times dc\_2(i) \times dc\_3(i) \qquad (2)$$

which heuristically are defined to tackle different aspects of the problems posed by using multiple ultrasonic sensors to measure distance. Each of them is established according to the following heuristic.

- A read value that presents a different value comparatively to the calculated by the system as expected should have a lower weight in the average. This degree of confidence 1, *dc_1*, shows the alignment of the sensor values with those predicted by a least-squares linear predictor/ estimator of the surface's inclination.

  The calculus of the surface slope is used to predict the surface value, using the past story The three last values of the average, *M*, are stored in a temporal window. If there is a big variation between the average value calculated using the measurements and the expected value, then the temporal window and the expected value should be reinitialised, because there has been a change in the surface characteristics.

- The read value should have less weight in the average calculation if, when compared to other sensors under the same circumstances, it presents a deviation greater then the one allowed. This sensor should have its rate of errors incremented. This deviation and the rate of errors determine the degree of confidence 2, *dc_2*. This *dc_2* expresses the agreement of each sensor measure with the measures of the other 3. The underlying assumption is that one is dealing with surfaces of smooth and small curvature.

The fact that a sensor is positioned near the boundary region introduces the degree of confidence 3, *dc_3*. This was specifically set-up to account for the detection of the part's edges. Its calculation uses a sigmoid function to approximate the probability distribution that the grip is pointing effectively to the part.

4. TESTS AND RESULTS

We used Matlab 4.2c to visualise the obtained surface. The algorithm we used to visualise the surface uses splines to construct the surface. Although this could be an advantage, for this specific case it isn't because the interpolation function opposes itself to big variations, trying to minimise the second derivative. Because of this the acquired surface presents a rounded aspect.

The experimental results of this process were obtained through the acquisition of a rectangular flat part with 20x20 cm, where the disposition of the surface relatively to the acquisition plan varied from 0° to 20°, as displayed in the figure. In a second experience we put another piece, 9 cm thick with 10 cm of height and 20 cm of width in front of the latter one, as showed in the figure 7.b). In this way, we obtained a piece with a relief in the order of 9 cm.

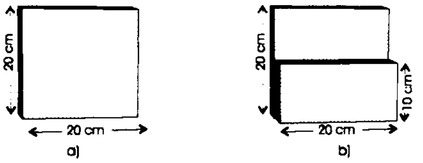


Fig. 7 - Pieces for acquisition **a)** Flat piece **b)** Piece with some relief

The figure 8 shows the surface obtained using an average weighted by degrees of confidence. The acquired surface was a flat surface with 20 x 20 cm, with an angle of 0° relatively to the plan of acquisition and 43 cm away from the ultrasonic sensors.

An analysis of the figure 8 shows that the shape of the object surface and its description were well retrieved.

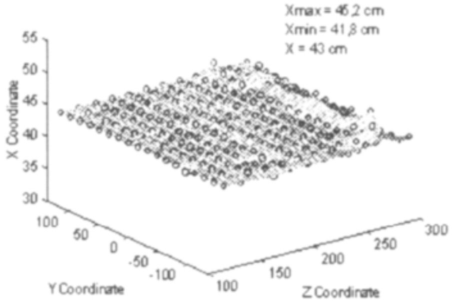


Fig. 8 - Experimental results for the map of a rectangular flat part with 0° of inclination

In the figure 9 we can see the obtained surface when the flat surface was positioned with 20° relatively to the scanning plan. As the maximum angle of incidence for these sensors is 25° the obtained results aren't unexpected.

The acquisition of a surface with 9 cm of relief shows that the fusion system minimises the slope of the surface, opposing to big slopes.

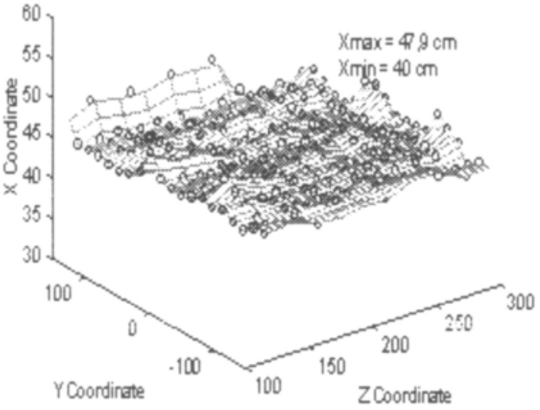


Fig. 9 - Acquisition of the flat surface with 20° of inclination relatively to the plan of acquisition

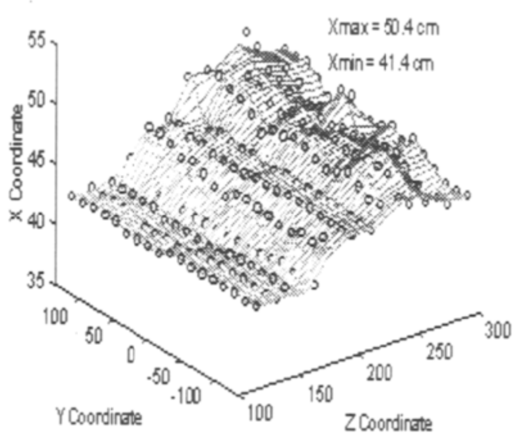


Fig. 10 - Acquisition of a surface with 9 cm of relief

5. CONCLUSIONS AND FUTURE WORK

A sensor system has been designed and constructed to acquire the profile of surfaces, based in a CCD camera for object boundary-determination and ultrasonic sensors for depth measurement. In order to reduce the measurement error resulting from the beam opening angle we covered the ultrasonic sensors with a tube of 20 cm and incremented the working frequency.

We used an average weighted by degrees of confidence for raw sensor data fusion, based on a heuristic set of rules. Practical experiments showed that the measurement errors when conventional ultrasonic sensors are used alone can be reduced using this multisensor system.

The experimental tests made with the new architecture for the PUMA 560 manipulator showed that it is possible to implement control actions more sophisticated comparatively to the old architecture.

The CAN network showed that it is highly reliable even when used under noisy environments. This feature is spreading CAN network in industrial environments.

This study demonstrates an encouraging approach to integrate several sensors at different levels. Surfaces more complexes, such as big inclinations relative to the plan ot acquisition or big variations of depth inside the object, have to be treated in other ways. In order to reduce the angle of inclination relative to the acquisition plan, the grip axis should be normal to the estimated surface curvature. In other situations other kind of more directional sensors, like infrared sensors, should be used and their data included in the sensor fusion.

There is a lot of future work to be done in this project, including improvements to the results already obtained. Current work is in progress for the adaptation of the depth measuring grid resolution to the surface curvature.

To minimise the movement duration of the manipulator, direct control of the joints is being considered, eventually with non-linear techniques. These sophisticated control actions are now possible because the new architecture provides all information necessary for these tasks. Regarding the geometry's acquisition time, high level techniques as clustering shapes and the correspondent class creation, as well as learning on geometry is being envisaged.

REFERENCES

1. A. Watt, *3D Computer Graphics*, Addison Wesley, 1993.
2. C. Santos, N. Monteiro, J. Fonseca, P. Garrido and C. Couto, "Control of a Robot Painting System Using the Multiresolution Architectural Principle", *in Proceedings of the IEEE International Symposium on Industrial Electronics - ISIE97*, Guimarães, Portugal, 7-11 July, 1997, pp. 672-677.
3. C. Santos, *Uma exploração do Modelo de Referência RCS e da Fusão Sensorial em Robótica*, Master Thesis, Guimarães, Universidade do Minho, October 1997
4. *CAN - Komponentem System Tools*, Steinbeis Transferzentrum Prozessautomatisierung , 1996
5. E. Piat and D. Meizel, " Range Data Fusion for occupancy grid building," in *Proceedings of the 4th International Symposium on Intelligent Robotic Systems '96*, pp. 353-360, Lisbon 1996.
6. Foley and Hughes, *3D Computer Graphics & Principle*, Addison Wesley, 1992.
7. G. Blais and M.D. Levine, "Registering Multiview Range Data to Create 3D Computer Objects". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17** (8) p.820-824.
8. H. F. Durrant-Whyte, "Consistent integration and propagation of disparate sensor observations". *Int. J.* Robot., Res., vol. 6, no. 3, pp. 3-24, 1987.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Features of the lateral active pneumatic suspensions in the ETR 470 high speed train

Massimo Sorli[a], Walter Franco[a], Stefano Mauro[a], Giuseppe Quaglia[a], Rocco Giuzio[b], Giuseppe Vernillo[b]

[a]Dept. of Mechanics - Politecnico di Torino - Corso Duca degli Abruzzi 24 - 10129 Torino -Italy
[b]Ricerca ed Innovazione Prodotto - FiatFerroviaria S.p.A. - Piazza Galateri 4 - 12038 Savigliano Italy

This work shows the methodology used to analyse the static and dynamic behaviour of the active lateral suspension on the so-called *Pendolino* train. The theoretical analysis of the system, comprehensive of the control part, allowed to define a simulation software capable of predicting the movements of the undercarriage on straights and on bends, using different types of suspension. Experiments carried out on prototype trains confirmed the validity of the modeling.

1. INTRODUCTION

This article concerns the research carried out on lateral active pneumatic suspension (LAPS) systems on high-speed trains. The research was carried out by the Dipartimento di Meccanica-Politecnico di Torino, and by Fiat Ferroviaria Spa, builder of the trains under study. The aim of the study is to describe the methodology used to analyse the static and dynamic behaviour of the lateral active suspension systems, fitted onto different and successive ETR 450, 460 and 470 versions of the *Pendolino* trains, and to show some of the experimental and numerical results obtained.

A high-speed train equipped with active suspension systems is a typical example of a mechatronic system, because an indispensable and precise integration of mechanical structures, servoactuators, sensors, and control is required, as these components are the elements which operate together to provide the vehicle's final performance. Rail high speed is usually obtained in two ways: either on purpose built tracks, or on existing tracks using vehicle capable of reaching higher speeds. The vehicles examined allow for effective operation in the second case. The suspensions represent a key element of the lateral dynamics and they have to satisfy two contrasting requirements: they have to guarantee a high level of comfort minimising body vibrations transmitted from the bogie, and they have to ensure that in all conditions the body to be safely centred with respect to the bogie. In order to conciliate between the different and opposing requirements, active devices can be added to the traditional passive suspension systems.

Many studies have been carried out in the field of active and semi-active suspension systems for railway vehicles. [1], [2] and [3] are significant and introductory examples of such studies. Some solutions for controlled suspension in vehicles are given in [4] and [5], while in [6] the performances of pneumatic suspension and passive suspension are compared. The study of the lateral dynamics of a railway vehicle with characteristics similar to the one under study was carried out in [7].

The article first describes the mechanical structure of the lateral pneumatic suspension on *Pendolino* trains. It then goes on to present the modeling of the servosystem with sensors, interfaces, actuators and control elements, connected to the model (kinematic and dynamic) of the mechanical part of the railway vehicle.

At the end of the article some of the experimental results obtained on a train prototype during tests on tracks are shown and compared to the numerical results of the simulator.

2. LATERAL ACTIVE SUSPENSION

Fig.1 shows the lateral suspension of the train studied. The diagram shows that between the bogie and the body there are not only springs but also pneumatic actuators. Furthermore, although not shown in the diagram, there are also vibration dampers.

The springs present low transversal stiffness in order to reduce vibrations transmitted to the body. When a high-speed train goes through a turning on tracks which are not specifically designed, a high lateral acceleration not counteracted by banked curving arises; this is denoted by A_{nc} (not compensated acceleration). Under these conditions the action used to balance the centrifugal force is produced by the pneumatic actuators. These therefore allow for the adequate centring of the body in relation to the bogie. The pneumatic plan provides for two actuators on each bogie. The connections and displacement are such that they guarantee a symmetrical configuration and hence an identical pushing effect on both sides of the body.

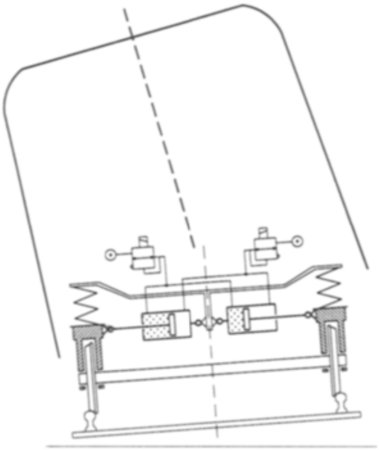


Figure 1. Lateral active suspension

The control of the force is guaranteed by the pressure in the chamber, supplied by means of two pressure proportional valves. The figure shows what happens on a left turning. In this case the proportional valve on the left is activated to produce a pressure rise in the front chamber of the left actuator and in the back chamber of the right actuator. This determines a force to the left which compensates the tendency of the body to exit the outer side of the bend. The figure also shows the relative rotation between the body and the bogie (rolling or tilting motion) which in this particular vehicle allows the passenger to be less affected by the centrifugal force.

The functional layout of the suspension system for each body is shown in fig.2.

The features that the control must provide can be seen as a trade-off between the following items:

a) robust curve recognition algorithm, able to avoid activation of commands due to track irregularities and points;
b) hard filtration of the input acceleration signals, considering both the introduced delay and the needed calculation time;
c) low calculation time of data, in order to identify also the entrance point of the curve. This objective has to be attained so that the action of the active servosystem is operative also in the initial phase of the curve, to avoid the suspension bumper from beating (in particular the front suspension);
d) bit numbers of A/D and D/A converters, according to the requested accuracy;
e) command synchronisation between the front and the rear bogie suspensions of the same body, according to the speed of train and to the dynamic performances of the control hardware;
f) efficient command synchronisation between successive bodies, considering acceleration sensors only placed on the front bogies of the first car body of the train, in order to guarantee the correspondence between the command start time and the body position on the trajectory.

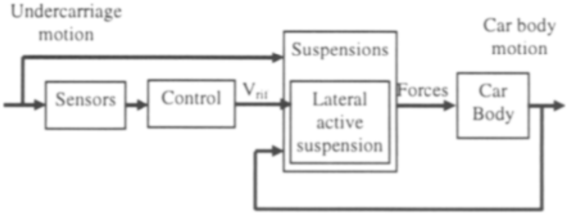


Figure 2. Layout of the suspension system

3. ANALYTICAL MODELING

To be able to assist the project phase, assess in advance the performance of the suspension, or make improvements with respects to existing situations, it is necessary to deeply analyse the problem, in order to obtain a mathematical model, to be validated with suitable experimental readings.

The simulator of the lateral dynamics of the vehicle, developed with Matlab-Simulink, is able to evaluate the rigid motion of the car body on a plane parallel to the tracks.

The model was conceived with the aim of reaching the following objectives:

1. the possibility of using inputs to the model both numerical (obtained from an analytical

computation of the trajectory), and experimental (acquired on location);

2. analysis travel both on straights and curves;
3. the possibility of simulating different types of passive and active lateral suspension, with relative control and regulation devices of the power;
4. the possibility of calculating the force on the traction lever of the car body;
5. obtaining sufficient accuracy on results relating to relative body/bogie movements, and to the lateral acceleration of the body, with a limited number of degrees of freedom.

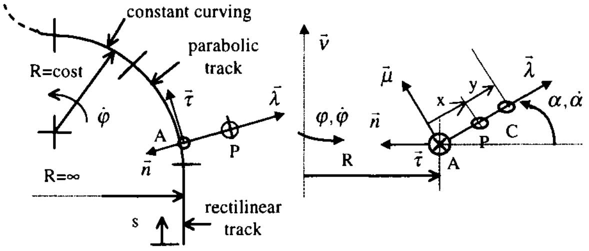
Assumed a defined body/bogie suspension, the lateral motion of the body is influenced by numerous factors, such as geometry of the tracks (curve radius and angle α of track superelevation), irregularity of the line, track/wheel contact, train speed and phase shift between actions on front and rear carriages.

The kinematic analysis conducted in [7] has attempted summarily to express the interaction between these factors. It demonstrates that the knowledge of A_{nc} is sufficient to define the dynamic behaviour of the train model with a satisfactory accuracy in relation to the desired objectives. In this analysis the carriages travel on tracks with sections having constant curving alternated with sections having variable curving, as indicated in fig.3a. We assume that the motion of the undercarriage takes place along the direction $\vec{\lambda}$ on a reference system $\vec{\lambda}, \vec{\tau}, \vec{\mu}$, where $\vec{\lambda}$ and $\vec{\tau}$ belong to the track plane, inclined by the angle α with respects to the horizontal (fig.3b). A, the origin of the axes, travels on the ideal trajectory and its position is defined by the curvilinear abscissa s. The angular velocity of the reference system is $\dot{\varphi}\vec{v} - \dot{\alpha}\vec{\tau}$. At any given point in time, P and C are the connection points to which the lateral suspension is constrained, belonging to the bogie and to the body respectively; x is the displacement $\overrightarrow{P-A}$, indicative of the movement of the undercarriage on the track and of the irregularities of the track itself, while y is the displacement $\overrightarrow{C-P}$.

Let the radius of curvature be R, the angular velocity $\dot{\varphi}$ (equal to $\dot{s}/R$), and the angular superelevation velocity $\dot{\alpha}$, the acceleration a_{PR} (measured with a monoaxial accelerometer placed on the bogie), considered positive with $\vec{\lambda}$, is approximable with [7]:

$$a_{PR} = \ddot{x} - x\dot{\alpha}^2 - x\dot{\varphi}^2 \cos^2 \alpha - \frac{\dot{s}^2}{R}\cos\alpha + g\sin\alpha \qquad (1)$$

while the acceleration of P is $a_{P\lambda} = a_{PR} - g\sin\alpha$. The measured acceleration a_{PR} is filtered by a low-pass filter, typically around 1 Hz, becoming A_{nc}.



3a) ground view 3b) rear view

Figure 3. Reference system of the bogie

The orientation of the body is defined by the joint reference system $\vec{i}, \vec{j}, \vec{k}$ centred at the centre of mass G as indicated in fig. 4; ϑ is the yawing angle. The angular speed of axes $\vec{i}, \vec{j}, \vec{k}$ is $\dot{\varphi}\vec{v} - \dot{\alpha}\vec{\tau} + \dot{\vartheta}\vec{k}$. C_1 and C_2 are suspensions constraints of the front and rear bogies respectively, being i_{p1} and i_{p2} their distances from G.

The accelerations of points C_m (m=1 front, m=2 rear) in the yawing plane $\vec{i}\vec{j}$ are given by:

$$\begin{aligned} \vec{a}_{Cm} = \vec{a}_{Gij} \pm i_{pm} \{ & [2\dot{\alpha}\dot{\varphi} sin\alpha sin^2 \vartheta - \dot{\varphi}^2 sin^2 \alpha sin\vartheta \cos\vartheta + \\ & + \dot{\alpha}^2 sin\vartheta \cos\vartheta - \ddot{\vartheta} - \ddot{\varphi}\cos\alpha]\vec{i} - [(\dot{\varphi} sin\alpha \cos\vartheta - \dot{\alpha} sin\vartheta)^2 + \\ & + (\dot{\vartheta} + \dot{\varphi}\cos\alpha)^2]\vec{j} \} \end{aligned} \qquad (2)$$

where $\vec{a}_{Gij}$ is the acceleration of the centre of gravity of the car body contained in the same plane and plus sign holds for m=1.

Equations (1) and (2) can be simplified by taking into account the effective values of the involved quantities, so that the terms $x\dot{\alpha}^2$ and $x\dot{\varphi}^2 \cos^2 \alpha$ in (1) can be neglected, and ϑ can be considered small ($\cos\vartheta \approx 1, sin\vartheta \approx \vartheta \to 0$) in (2).

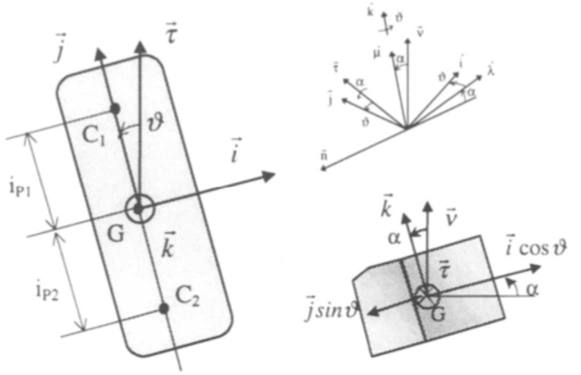


Figure 4. Reference system

By indicating the component of the absolute angular acceleration of the body in $\vec{k}$ direction with $\ddot{\gamma} = \dot{\varphi}\cos\alpha + \ddot{\vartheta}$, under the previous assumptions, the body/bogie front and rear relative accelerations in the direction $\vec{\lambda} \equiv \vec{i}$ can be expressed as:

$$\ddot{y}_m = a_{C\lambda m} - a_{P\lambda m} = a_{G\lambda} \mp i_{Pm}\ddot{\gamma} - a_{PRm} + gsin\alpha \tag{3}$$

where minus sign holds for m=1. The equilibrium of the body on the lateral plane, being M the mass, I the yawing moment of inertia, V_1 and V_2 the exchanged forces in the secondary suspension, is given by:

$$\begin{cases} a_{G\lambda} + gsin\alpha = -(V_1 + V_2)/M \\ \ddot{\gamma} = (V_1 i_{P1} - V_2 i_{P2})/I \end{cases} \tag{4}$$

having considered negligible the term $\dot{\varphi}\dot{\alpha}sin\alpha$ in the angular momentum derivative, simplified as $I\ddot{\gamma}\vec{k}$ [7].

V_1 and V_2 are defined as:

$$V_m = V_{am} + \beta_m \dot{y}_m + k_m y_m \tag{5}$$

where for m=1,2, V_{am} is the thrust force of the pneumatic servoactuator, β_m and k_m one, respectively, the lateral damping coefficient and the lateral stiffness of the suspension's passive elements. The only input is the acceleration of the front bogie. Assuming that the acceleration of the rear bogie is equal to the front one apart from a time shift dependent on the speed of the train, the system of equations (1)-(5) can be numerically integrated, to determine all the kinematic values. In particular, the contribution of the active servoactuators can be computed. Their forces are a function of A_{nc}.

4. SOFTWARE SIMULATOR

Many interacting subsystems are part of the physical system of the high-speed train. In order to highlight the performances of the active lateral suspension, or the effect of force V_{am} in relation to the control mode, the more relevant subsystems are the car body, the suspension and the control. Fig.5 shows these elements and their functional connection to each other.

The excitation of the system is caused by the motion of the two bogies. Under the previous hypothesis, this motion is entirely defined by the acceleration measured on the front bogie a_{PR1} which constitutes the input of the simulator. This is used both for the command of the pneumatic servoactuators, based on proportional algorithm implemented in the 'Control' block, and to determine the car body/ bogie relative motion. In the 'Mechanical input' block a_{PR1} is filtered and phase shifted depending on the train speed to produce the signals a_{PR1F} and a_{PR2F}, the entrances to the 'Kinematics' block in which equations (3) are introduced.

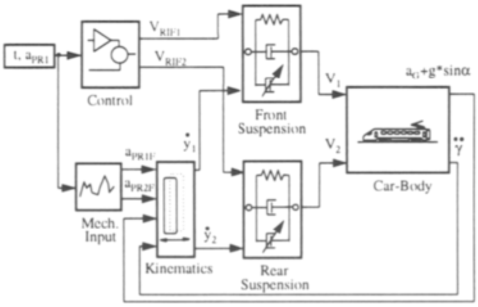


Figure 5. Lateral dynamic simulator

The 'Front suspension' and 'Rear suspension' blocks determine the lateral force V_m on the car body as in equation (5), based on the relative motion, and on the reference signals V_{RIF1} and V_{RIF2} for the servoactuators outputs of the control. The law of motion of the car body is obtained from the 'Car body' block, as in the system from equation (4).

As shown in fig.6, which visualises the content of the 'Front suspension' block, the force V_1 is the sum of the passive force V_{p1}, and of the active force V_{a1} generated by the pneumatic servoactuator.

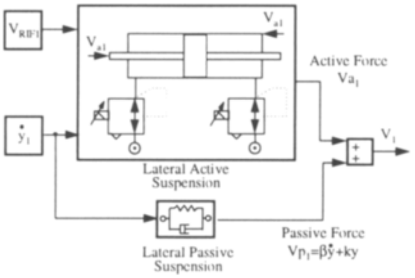


Figure 6. Front Suspension

Considering the layout of fig.1, the two asymmetric pneumatic actuators are equivalent to one symmetric actuator, which rod is constrained to the bogie while the barrel is constrained to the body.

The 'Lateral active suspension' block is blown up in fig.7. The nonlinear model of the pressure proportional valves, which describes its static and dynamic characteristics as a function of the reference V_{RIF} and of the mass flow rate G, determines the output pressure ($press_{1L}$, $press_{1R}$, where subscript L=left and R=right). Having defined the flow rates G, calculated in the resistive models of the pipe, and the relative speed $\dot{y}$, in the 'cylinder' block the thrust force V_{a1} is calculated using continuity and equilibrium equations.

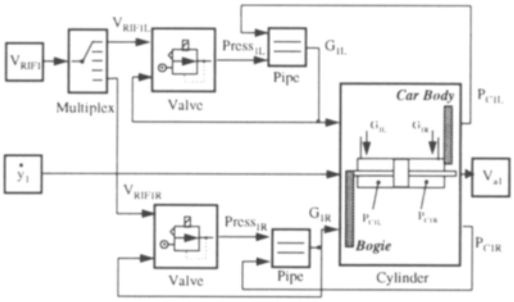


Figure 7. Lateral active suspension

5. EXPERIMENTAL RESULTS AND IDENTIFICATION OF THE MODEL

The identification of the model parameters was carried out on the basis of several on-track tests on a train prototype which was set up with the appropriate instrumentation.

In fig. 8a) a typical A_{nc} signal is shown: it was registered in a particularly difficult section of track, at a speed of 128 km/h. Negative A_{nc} corresponds to travel over a right curve. The considered section first has a slight right curve, with continually variable radius, followed by a left curve with a small radius of curvature, then another right curve followed by one to the left, with no straights in between. The A_{nc} grows in modulus at the entrance of the curve, remains approximately constant in sections with constant radius and banking, and decreases to null once the curve is finished.

In fig. 8b) simulated pressures inside the chambers of the cylinders are shown.

The suspension systems, depending on the command received by the control and by the car body/bogie relative motion, generate a lateral force on the car body itself, with the aim of keeping it centred during curve travel.

Fig.8c) shows simulated and experimental trends of the front relative displacement y_1. Right displacement of the car body is positive.

Rear relative displacements are shown in fig.8d).

The simulator, upon receiving as an entrance the A_{nc} signal obtained during on-track tests, reproduces both the qualitative trend of the car body's movements, and their maximum values. This result confirms the correctness of the parameters used in the 'Control', 'Suspensions' and ' Car body' blocks.

Therefore the model can predict the behaviour of the train's active lateral suspension as a function of the A_{nc} signal, whether this is registered on line or predetermined theoretically. It is then possible to modify the control strategy of the servoactuators in order to optimise the suspension's performances.

The shapes of the proportional valves output pressures are well overlapped to the A_{nc} signal, according to the proportional control strategy. Because of the small cylinder stroke and the low disturbance speed, also the thrust forces are on phase with A_{nc} input. In other word, the bandwidth of this type of pneumatic servosystem fits with the assumed control.

The presence of relevant relative displacements in correspondence with high A_{nc} gradient suggests that is possible to operate on the control side, both on curve recognition and on supply procedures of the valve reference signals.

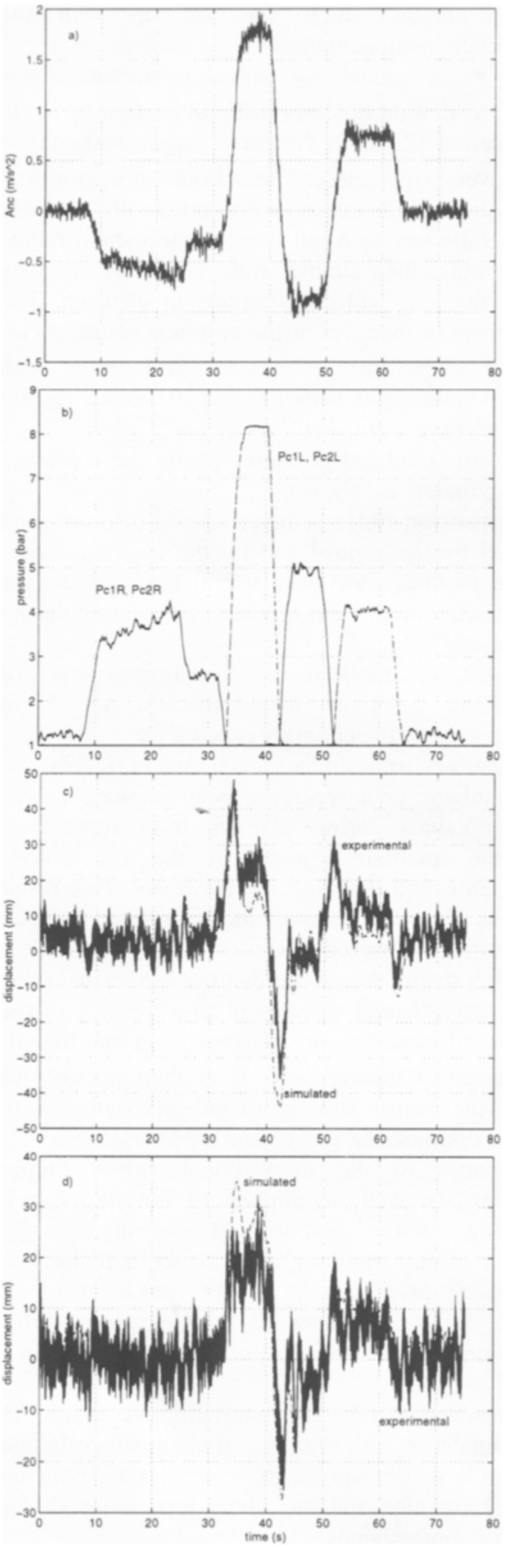


Figure 8. Experimental and numerical results

6. CONCLUSIONS

This article briefly summarises the more relevant activities that were carried out during a research concerning the performances of high-speed trains. The simulator proposed is able to provide the kinematic variables (relative displacements, speed, accelerations) the forces exchanged between bogie and car (passive part and active part), the pressures, the flow rates and the references of the pneumatic components. One particular aspect of the model is the possibility of operating without a direct knowledge of the track's angular superelevation.

This research first of all enables a sufficiently accurate identification of the existing situation; moreover it gives full understanding of the influence of the characteristics of both the pneumatic component, and of the control structure on the system performances. Finally, experience lead to the proposal of modifications, particularly aimed at control and sensorisation, with the objective of increasing the system's readiness.

REFERENCES

1. Jiudai K. et al.:»Fundamental study on semiactively controlled pneumatic servo suspension for rail cars» Proc. Joint IEEE-ASME Railroad Conference, Atlanta, 1981, pp 1-7.
2. Okamoto I. et al.: «An active suspension system for railroad passenger cars», Proc. Joint IEEE-ASME Railroad Conference, Toronto, 1987, pp.141-146.
3. Cho D., Hedrick J.K.: «Pneumatic actuators for vehicle active suspension applications», Journal of Dynamic Systems, Measurement and Control, vol. 107, marzo 1985, pp.67-72.
4. Ferraresi C., Quaglia G., Sorli M.: «Force control laws for semi-active vehicular suspensions», European Journal of Mechanical Engineering, vol. 42 n°3, pp. 303-308.
5. Sorli M., Quaglia G., Mauro S.: «Semi-active hydropneumatic suspension» Proc. Mechatronic Conference, 27th Int. Symp. on Automotive Tecnology and Automation ISATA, Aachen, 1994, pp.173-180.
6. Belforte G., Ferraresi C., Sorli M. «Design analysis and dynamic behaviour of a pneumatic suspension», Proc. third JHPS Int. Symp on Fluid Power, Yokoama, Japan, 1996, pp.133-138.
7. Quaglia G., Sorli M.:"Analisi della dinamica laterale di un veicolo ferroviario" Proc. XIII National Congress AIMETA, Siena, Italy, 1997, pp.303-308

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronics in automating shunting processes in marshalling yards

I. Jenayeh, Ch. Müller, N. Kositza, M. Enning and H. Rake

Institute of Automatic Control,
Aachen University of Technology, 52056 Aachen, Germany
Phone: ++49-241-807481, Fax: ++49-241-8888296,
E-mail: je@irt.rwth-aachen.de

The German Railway (*Deutsche Bahn AG*) intends to introduce an automatic coupler for freight cars to increase the efficiency of the marshalling process and to humanize it. To speed up the train-handling process in the approach zone and the train-formation process in the distribution yard an actuator is needed to automate the unlocking of these couplers while freight cars are moving above the hump with a speed of up to 3 m/s. For these purposes a new manipulator and an extension of the automatic coupler [1] have been developed and realized at the Institute of Automatic Control in Aachen, Germany, and are presented from a mechatronic point of view in this paper.

1. INTRODUCTION

To improve the competitiveness with road transportation (heavy goods vehicles) the German Railway (DB AG) has been striving for several years to decrease the transport time of freight and at the same time to increase the volume of freight. To achieve this, several handling activities in the marshalling yards (Fig. 1), which are considered to be the bottleneck in rail transport, have been automated.

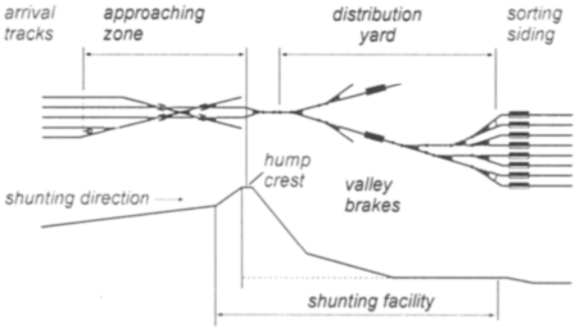


Figure 1: Principle of a marshalling yard

Only the handling of freight cars in the approaching zone as well as in the distribution yard has still to be done manually while cars are standing. Therefore, manipulating the conventional screw couplings is slow, labor-intensive, time-consuming and a considerable source of accidents. For these reasons several attempts have been made to develop an automatic train coupler. In a collaboration with Knorr Bremse of Germany, the DB AG has devised a new automatic coupler (Z-AK), which transfers only pulling forces while the bumpers remain.

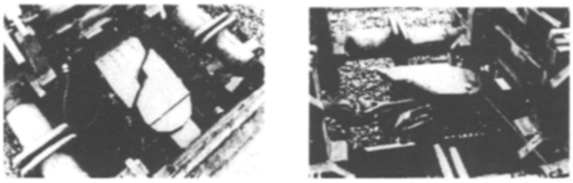


Figure 2: The automatic coupler (Z-AK)

This coupling (Z-AK, Fig. 2) allows in addition to the (automatic) mechanical connection also the pneumatic and electrical one. Furthermore, this coupling can be operated from the track side by pulling a hand lever located under the side buffer. Thus, the uncoupling is the only manual operation in the marshalling yard. Fig. 2 shows also the possibility of mixed coupling, which is a very important aspect to consider in the design of this coupling. This mixed coupling allows a longterm introduction and the conveyance of foreign freight cars which don't have these automatic couplers.

To bridge the resulting gap in automation of marshalling yards the Institute of Automatic Control of Aachen University of Technology in Germany has developed a machine, which should be cheap, reliable and easy to install at the hump. After developing a mechanical extension of the automatic coupling, the latter became well-suited for automatic manipulation, and it became possible to develop a manipulator which can be built next to the track and operate automatic couplers of freight cars moving with a speed of up to 3 m/s.

The next sections will reflect slightly on this mechatronic system. The combination of the mechanical extension and manipulator with the control system based on a high-performance processor will be highlighted in this paper.

2. MECHANICAL PART OF THE AUTOMATION

2.1. Principle

As delivered by Knorr, unlocking the autocoupler has to be done manually by pulling a lever from the track side. The shape of this lever and its motion, which has to be performed, are not well-suited for automatic handling. The construction of the lever was based on aspects of manual handling without considering an automatic handling. A first attempt to automate this uncoupling process led to a very sophisticated machine, which seems to be too expensive and not robust enough for the operating conditions at the hump [2,3].

Consequently, an extension to the Knorr auto-coupler has been developed by the Institute of Automatic Control. This mechanical extension consists of a kind of gear that transforms a simple swivel motion into the needed pulling motion (Fig. 3). The swivel motion is realized by just blocking the lever while the wagons are moving.

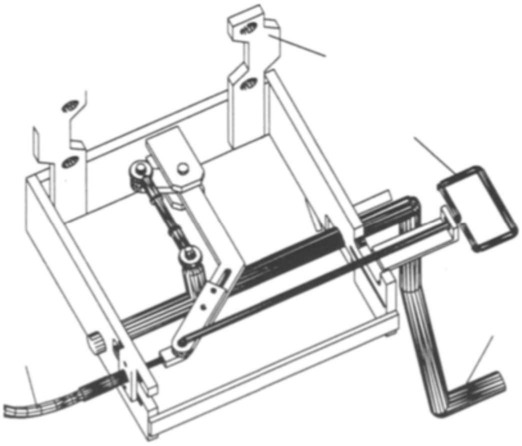


Figure 3: Structure of the extension

The pulling motion is transmitted by a Bowden cable to the coupler (Fig. 4) and is as long as the motion of the hand lever. The Bowden cable enables the extension to be positioned at any appropriate location. The space just below a side buffer has been found as an optimal position, since this area has to be held free according to international conventions (UIC leaflet 535-2 V).

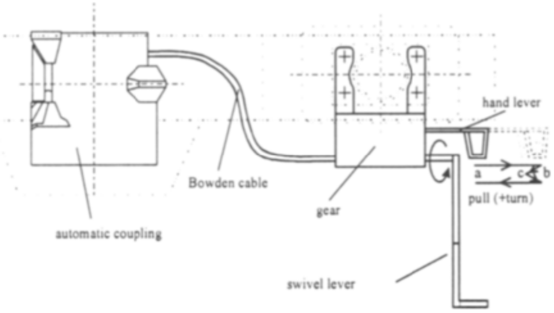


Figure 4: The operating device

Automating the handling of the coupling requires a manipulator which is able to automatically detect the moving cutting-off point, to localize it and to handle the swivel lever. Figure 5 outlines the principle of the automatic uncoupler.

As shown the swivel lever is blocked by a stationary manipulator while the freight cars are shunted. When the swivel lever reaches its final position the blocking force increases rapidly and thus acts as a trigger signal and the lever hand (manipulator) is retracted. To minimize the moment of gravity in the neutral position of the swivel lever and to maximize it for falling back a bent shape of the lever has been developed.

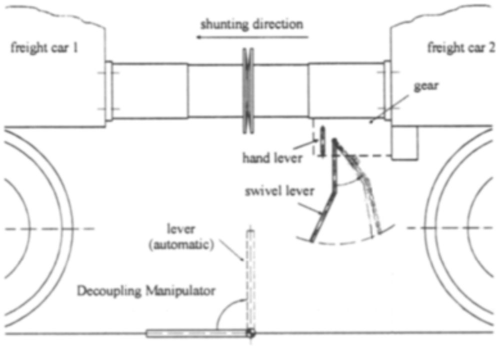


Figure 5: Basic assembly of the new uncoupling manipulator

The decoupling manipulator, a simple swivel arm, is raised from the resting position, parallel to the rails, into the vertical at the right moment. To guarantee the safety of the manipulator, this motion should be opposite to the shunting motion. Thus in case of a wrong detection of the cutting-off point or a collision with parts of the freight cars the lever will be pushed back automatically to its resting position [4].

2.2. Realization

Because of high shunting speeds and high manipulating forces several requirements in the design of the manipulator and its control system had to be taken into consideration. At worst, a swiveling time (from the resting position to the vertical one) of about 100 ms is required. To achieve this, an appropriate maintenance-free AC-servo drive was chosen. Furthermore, high torques which are needed during the manipulation (up to 300 Nm), can be obtained by a gear mounted to the drive to a conventional motor.

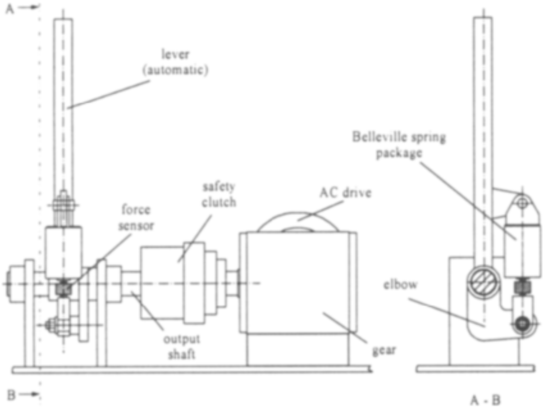


Figure 6: Structure of the manipulator

Figure 6 shows the structure of the manipulator. The drive and the gear are connected to the swiveling device by means of a safety clutch, which compensates any misalignment between the output and the gear shaft. This safety clutch should also disconnect the drive part from the swiveling one in case of overload. The lever is connected indirectly to the output shaft by means of a Belleville spring package and an elbow. The spring package damps force peaks and makes more time available for the retracting motion of the drive. Furthermore a push/pull force sensor was integrated to sense the force development and to determine the right retracting moment.

To minimize the mass moment of inertia the spring package and the force sensor are fastened as near as possible to the swiveling axis by means of the elbow. Figure 7 shows a photo of the realized uncoupling manipulator.

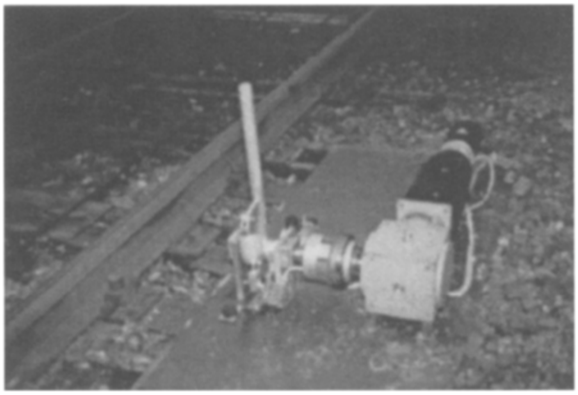


Figure 7: Realized uncoupling manipulator

3. ELECTRONIC PART OF THE AUTOMATION

3.1. Principle

After presenting the mechanical part of the automatic uncoupling process, an overview of the electronic structure and the implemented uncoupling strategy will be given in this section.

For a safe and reliable uncoupling process several sensors are necessary. By means of an encoder the position of the manipulator lever is detected. From the force sensor, the uncoupling force and the state of the uncoupling process (blocked, locked, opened, etc.) can be determined and monitored.

Additionally, a track switch and - if required - a radar sensor is used to localize the cutting-off point and to determine the right moment to swivel the lever hand. The track switch detects the approaching wheels. The wheels are counted and then compared to a computerized list (a-priori knowledge) which includes all the data about the shunted cars. The speed of the wagons is determined either by the track switch (instantaneously) or the radar sensor (continuously). Figure 8 gives an overview of the structure of the uncoupler plant.

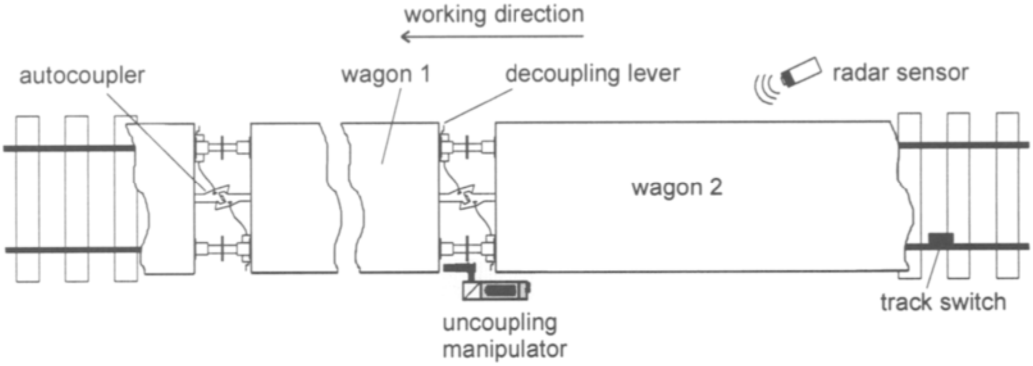


Figure 8: Automatic uncoupler plant

The estimated position of the coupling to be operated is calculated in real time. As soon as the cutting-off moment has been reached the lever must be swiveled within 100 ms from the resting position to the vertical one. During the following waiting phase the lever is kept in the vertical position. The waiting phase ends when the manipulator lever engages with the swivel lever.

From this moment the measured blocking force includes information about the state of decoupling. A first peak arises from the collision of the two levers. The next two peaks result from the process of the coupler opening. Finally the swivel lever reaches its final position and the force increases immediately. After reaching a preset threshold value of this uncoupling force the manipulator lever is retracted as fast as possible.

3.2. Realization

A high-performance-processor based system is used for all controlling purposes and consists of a Personal Computer (PC) and a digital signal processor (DSP)-based system. The PC operates as a monitoring and control system, the DSP works off the control operations.

Additionally the DSP-system offers a graphic-oriented programming interface to implement the controller software and is specially designed for rapid prototyping works. Therefore, the expenditure for the software development is kept very small. Figure 9 illustrates the control system structure.

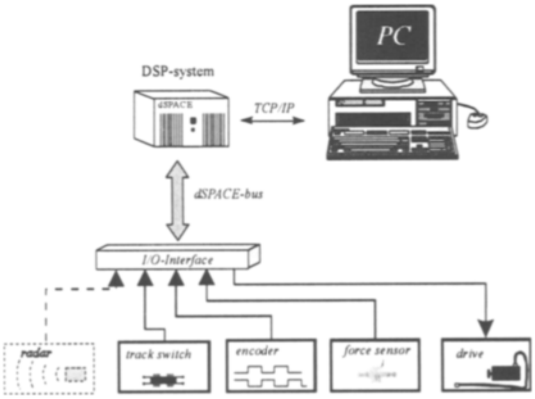


Figure 9: Control system structure

All sensors and the chosen electrical drive are connected to the control system to automate the process. In addition to the reasons mentioned above, this kind of drive offers flexibility of design and the possibility to test different strategies. This is reached by means of quick and easy adjustments in the controller software (e.g. compliance of the manipulator, force limits, time behavior), instead of time-consuming constructional solutions.

The uncoupling process can be completely determined and monitored. The control task consists of the time optimal positioning of the swivel lever of the manipulator, keeping it at the desired vertical position, guaranteeing good robustness against disturbances, especially the manipulating forces and retracting it after reaching the preset threshold.

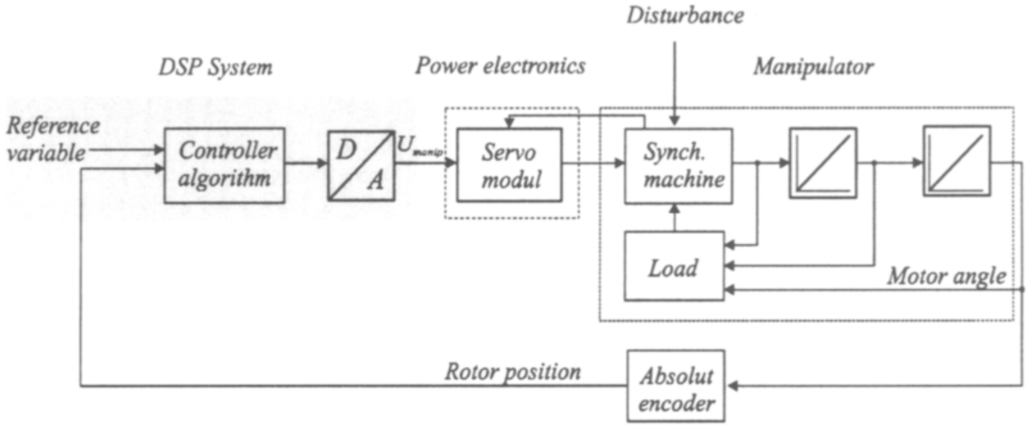

Disturbance
DSP System
Power electronics
Manipulator
Reference variable
Controller algorithm
D
A
U_manip
Servo modul
Synch. machine
Load
Motor angle
Rotor position
Absolut encoder

Figure 10: Block diagram of the control loop

To design a controller, a nonlinear mathematical model of the controlled system has been built up and implemented with MATLAB/SIMULINK for simulation purposes. Figure 10 illustrates the block diagram.

Based on this model a position control strategy has been developed, including a time-optimal two position controller and a linear output feedback controller [5]. The first one is designed to bring the lever next to the desired reference position as fast as possible. The second controller is designed to correct small position deviations and of course to compensate disturbances caused by the handling process.

4. REALIZATION OF THE PROTOTYPE

From a mechatronical point of view the synergy between the mechanical, electronic and computational design can been seen clearly. The manipulator prototype has been realized in Aachen, Germany, where trials under hard circumstances have shown the technical feasibilty of the new concept. Figure 11 to 13 show the manipulator during different states of the decoupling process; Figure 11 illustrates the rising of the manipulator arm, Figure 12 the unlocking of the coupling and Figure 13 shows the manipulator after the arm is retracted.

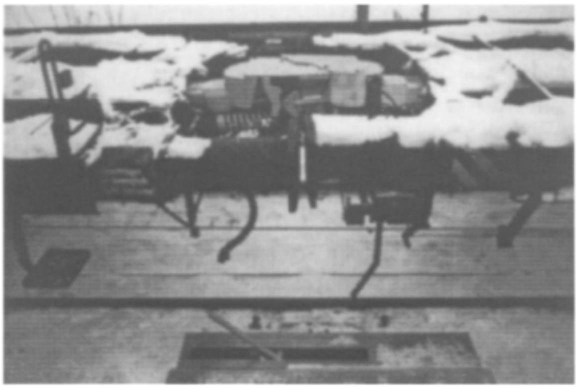


Figure 11: Rising of the manipulator arm

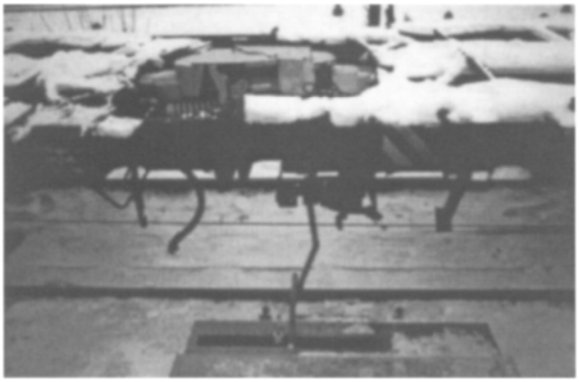


Figure 12: Manipulator while unlocking automatically the coupling

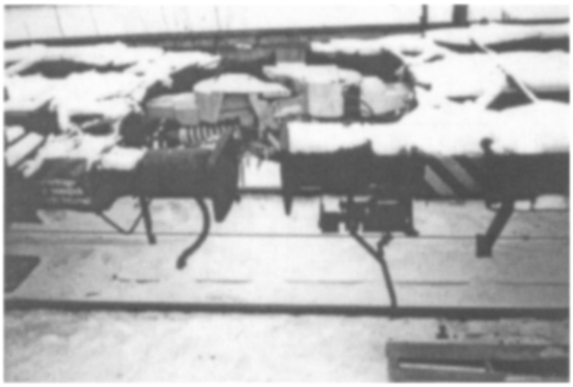


Figure 13: Manipulator after the retraction of the arm

REFERENCES

1. Institute of Automatic Control in Aachen and DB. AG (1996). *Einrichtung zum Entkuppeln von Schienenfahrzeugen mit automatischer Zug-Kupplung.* Patentanmeldung Nr. 196 37 546.0, September 1996.
2. H. Rake, M. Enning, J. Kurth and W. Schröder (1994). *Automatic uncoupler completes automation at the hump*, Railway Gazette International, June 1996, pp. 371-372.
3. W. Schröder, I. Jenayeh und H. Rake (1996*). Automatische Entkupplung von Güterwagen*, GMA-Kongress 1996, VDI-Berichte 1282, pp. 427-436, Düsseldorf.
4. I. Jenayeh, Ch. Müller, M. Enning, H. Rake and J. Peter (1997). *Development and Realization of a Manipulator for Automatic Decoupling of Freight Cars in Marshalling Yards*, World Congress on Railway Research WCRR '97, Volume E, November 1997
5. I. Jenayeh, Ch. Müller, N. Kositza, M. Enning and H. Rake (1998). *A new approach for automating shunting processes in Marshalling Yards*, Computational Engineering in Systems Applications IMACS Multiconference '98, April 1998

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

The Development of a Hardware-in-the-Loop Simulation of a Pump Storage Hydro Power Station

Sa'ad Mansoor[a], David Bradley[b], Capell Aris[c] and Glyn Jones[c]

[a]SEECS, University of Wales, Dean Street, Bangor, Gwynedd LL57 1UT, UK

[b]School of Engineering, University of Abertay; Dundee, Bell Street, Dundee, DD1 1HG, UK

[c]First Hydro Company, Dinorwig Power Station, Gwynedd LL55 4TY, UK

The operational flexibility of pump storage hydro-power stations means that they are often used by power system managers as their primary means of frequency control. However, as the operational demands on the power system, and hence on station operation, increase there is a need to better understand the behaviour of the station and in particular the operation of the machine governors. The paper therefore reports on the development of a generic model of a high head pump storage hydro-power station intended to form the basis of a hardware-in-the-loop simulation in which the real governor is used to control the system model.

1. INTRODUCTION

Unlike base load stations, pump storage hydro power stations are designed to be very flexible in operation and are able to pick up or shed load very rapidly, suiting them to system frequency control and load management. Such stations are complex, integrated systems operating according to mechatronic principles and involve the integration of a wide variety of sub-systems ranging from water management to the control and operation of the turbine/generator sets themselves. In the case of the Dinorwig and Festiniog power stations operated by First Hydro this integration is achieved by means of an integrated network deploying large numbers of PLCs with Festiniog being capable of being operated remotely from the control room at Dinorwig.

As increasing demands are placed by system managers upon the operation of pump storage hydro-power stations in terms of their ability to respond to system changes, there is a need to better define the station operating parameters in order to optimise performance under a wide range of system conditions. Of particular importance in this respect is the performance of the machine governors which regulate the power produced by the individual machines in response to changes in system frequency.

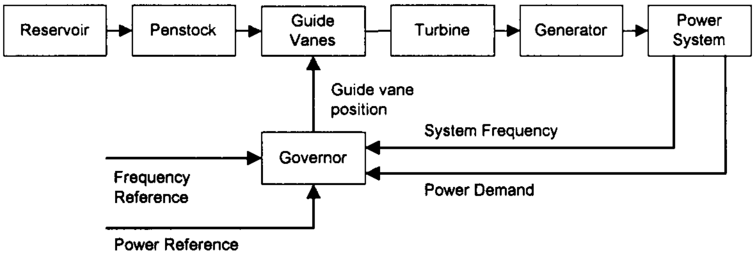


Figure 1: System block diagram

Referring to figure 1, the governor receives information on the system frequency and power demand and acts to regulate the flow of water through the penstock by control of the turbine guide vanes in order to restore the frequency to its set value. The governor settings are a major factor in determining the response of the plant and the determination of the optimum values of gain is therefore an important factor in improving overall performance.

However, the nature of the power station and the associated costs, and indeed risks, are such that it is generally impractical to carry out trials at the necessary level of detail without a significant possibility of damage to the power station and its plant. In order to provide the necessary basis for understanding governor, and hence system, behaviour a model of the power station is being developed which can be used as the basis of a 'hardware-in-the-loop' simulation in which the simulated governor is ultimately to be replaced by the real governor which will then used to control the simulated machines to enable the optimum settings to be determined for different operating regimes.

2. GOVERNOR OPERATION

The machine governor has the general configuration shown in figure 2 and comprises two main control loops as follows:

1. A frequency control loop which has the frequency error as input and the output of which is used to set the desired guide vane position. A dead band, typically of 0.05 Hz, would be included in this loop to limit the response and prevent hunting.
2. A power control loop in which the power demand is operated on by the droop gain α and is used to create a compensating signal.

In addition, the power reference signal is used to inject a feed-forward supervisory signal which is used to set the reference position for guide vane opening. This enables a rapid initial response to be achieved after which the governor acts to trim the guide vane position to obtain the required output power. By this means a response time from zero to fully opened (1 p.u.) is 28 seconds in normal operation and 8 seconds under emergency conditions can be achieved by Dinorwig.

The power reference signal is also used by the operators to define the operating point for machine when used in the part-load response mode for frequency regulation and control.

3. HYDRAULIC SYSTEM

As pump storage hydro-power stations typically operate with a high head, any model of the hydraulic system must take account of transients effects such as the pressure waves and water hammer effects resulting from operation of the guide vanes to

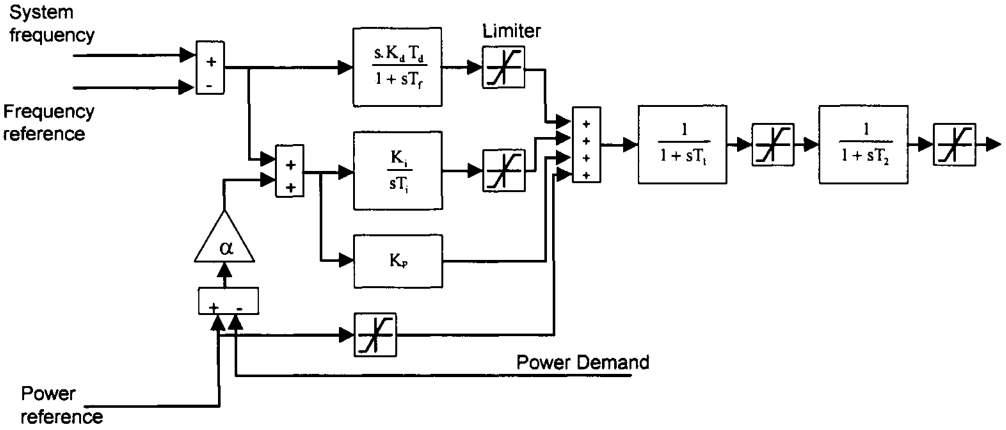


Figure 2: Governor control loops

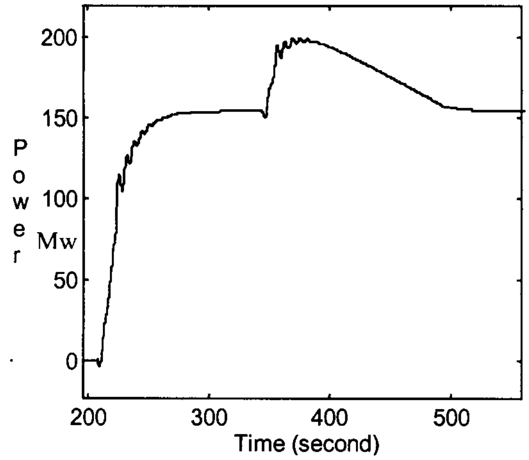


Figure 3. Response of pump storage station to a change in demand load

regulate the flow of water. This penstock is therefore modelled as a transmission line terminated in an open circuit at the turbine and a short circuit at the reservoir. For a large reservoir, the incremental head and flow at the turbine inlet can be represented by the relationship of equation 2 [1-3].

$$H(s)/Q(s) = -T_w/T_e \tanh(T_e s + F) \qquad 2$$

where $T_w = \sum L.q_{base}/h_{base}.g.A$ and is referred to as the water time constant; T_e is the wave transit time and F the coefficient of friction in the penstock.

Using the Maclaurin series approximation for tanh, the relationship between machine torque and guide vane position is given by equation 3:

$$\frac{M(s)}{Z(s)} = \frac{1 - sT_W + 0.5s^2T_e^2 - 0.167s^3T_WT_e^2 + 0.04167s^4T_e^4}{1 + 0.5T_W + 0.5s^2T_e^2 + 0.0834s^3T_WT_e^2 + 0.4167s^4T_e^4} \qquad 3$$

This is then the representation that is used in the simulation.

4. SIMULATION AND MODELLING

The normal mode of machine operation for frequency regulation and control is the part-load response mode. In this mode of operation the machine is typically loaded by the operator to 50% of full load (150 MW) using the governor power reference input. When the system is subject to an increase in demand load this is reflected as a change in system frequency and the governor will respond by increasing the power generated by the controlled machine to compensate and restore the frequency to its defined value. In a real system, the frequency control station will not be required to carry the full added load for significant periods of time and as other, slower responding, plant comes on-line it will then shed load to return to its defined operating point [4].

The simulation of the operation of Dinorwig pump storage hydro-power station in part load response mode was simulated using MatLAB/Simulink. Referring to figure 3, following synchronisation with the grid at 220 seconds into the simulation the generator is loaded to its normal operating point of 150 MW. This represents the part-load operating mode for the station.

A step load of 50 MW is then applied to the system at around 350 seconds into the simulation and the station responds with a corresponding increase in generated power over a period of the order of 20 seconds. As slower responding plant then comes on line and picks up the additional load, the generator falls back to its 150 MW set point. In this case of figure 3 this condition is reached at around 500 seconds into the simulation.

The nature of the governor response obtained will be dependent on the governor settings, the current grid loading, the nature of the demand load and the response of remote plant, all of which can be included as part of the simulation.

4.1. Relationship between governor settings and grid stiffness

The condition of the grid is a major factor in determining the governor settings required for optimal operation. By using the model the influence of the grid on governor behaviour can be established and the limiting conditions for governor gains

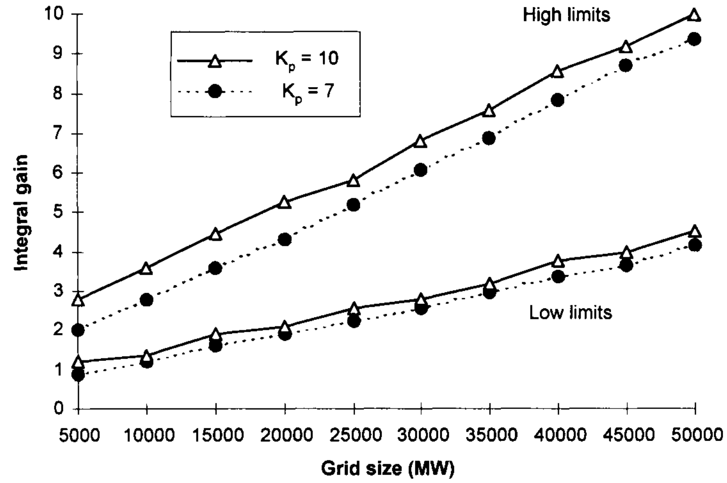


Figure 4. Relationship between grid size and governor settings

considered. Figure 4 shows the relationship between the governor gains and the grid stiffness in which the low limits define a slow stable response with minimal oscillation and the high limits represent a faster response with acceptable oscillation.

Figure 5 then shows the effect on the frequency response obtained for various sizes of grid with the governor settings of table 1.

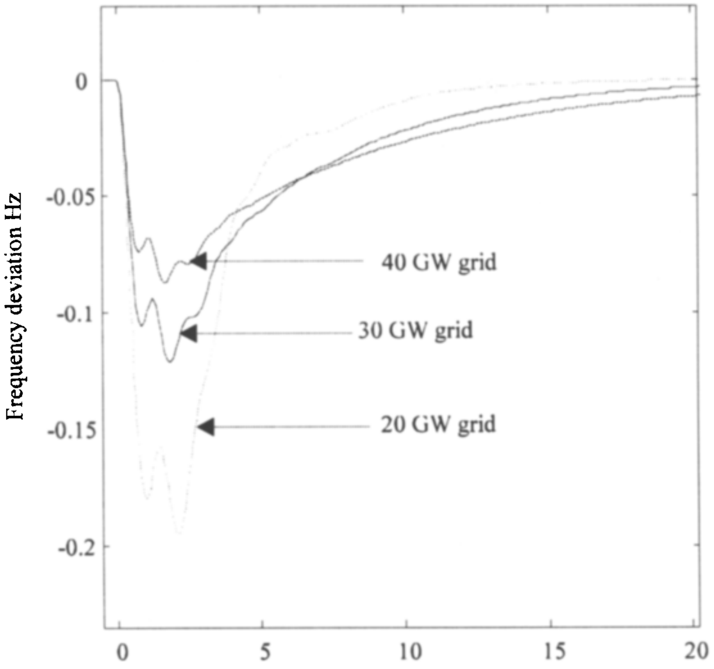


Figure 5. Frequency deviation

Table 1
Governor settings for Fig. 5

K_p	K_i	K_d
10	7	4

4.2. Multi-machine operation

The models reported thus far have assumed a single machine connected to the grid and supplied from a single penstock. At Dinorwig machines are supplied from a common tunnel from which individual penstocks feed the individual machines. This means that there will in practice be significant interaction between machines both in they hydraulic system and the electrical system.

The basic model was therefore revised to accommodate a multi-machine system supplied from a common tunnel with individual penstocks with appropriate interconnection on the electrical side.

This model was then tested for a variety of different conditions and the response obtained compared with that obtained from the Dinorwig machines. A particular point of interest was to investigate the effect that a slight 'detuning' of individual governors with respect to each other would have on machine and station performance.

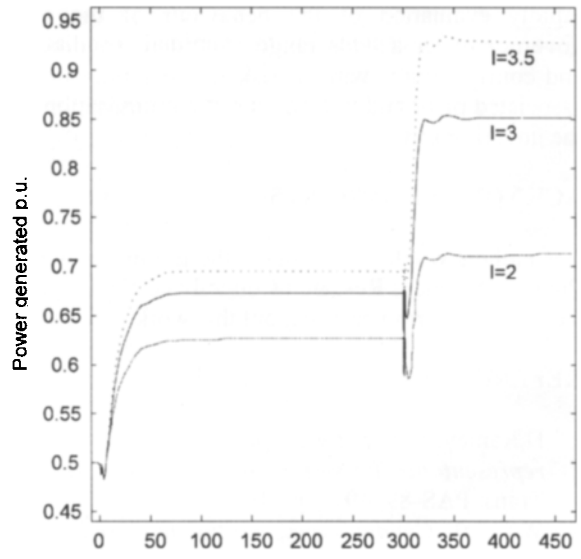


Figure 6. Simulation of a three machine system with different P-I-D parameters

This is reflected in figure 6 in which the load sharing performance of three machines whose integral gains are set to slightly different values is shown.

The understanding of this particular condition is important since the replacement of the previous analogue governors by digital governors has eliminated any such differences that would have existed previously and which may, under some circumstances, have been beneficial in terms of station operation. Figure 7 shows the non-linear hydraulic model used for the multi-machine system.

5. HARDWARE-IN-THE-LOOP SIMULATION

The simulation has been designed and developed such that the governor model receives the same input signals and provides the same output signals as the real governor. In the next stage of the work, the MatLAB Real-Time Systems toolbox will be used to allow the replacement of the model governor with a real digital governor from one of the Dinorwig machines which will then be used to control the system model in response to simulated changes in system conditions.

This will then enable the response of the real

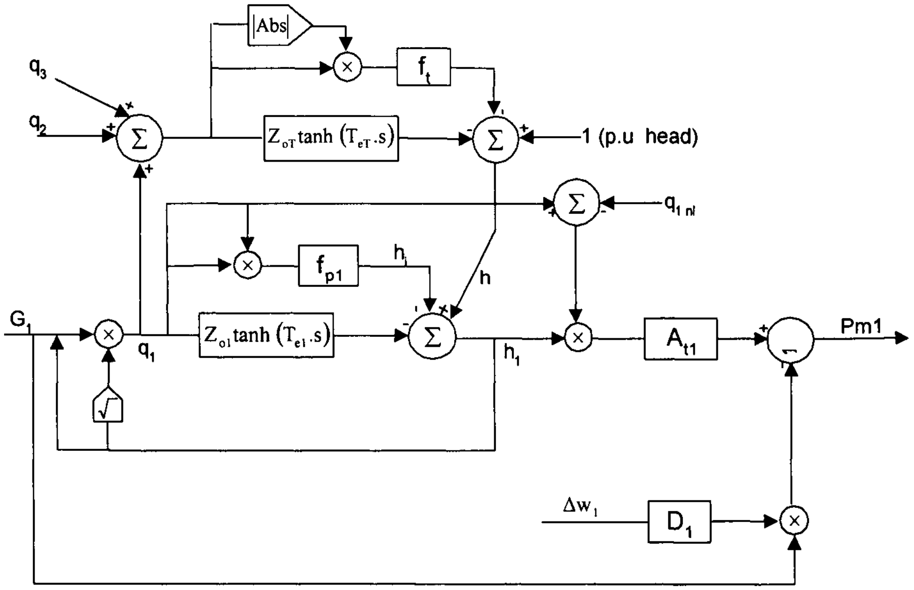


Figure 7. Non-linear model of multiple penstocks & turbines supplied from a common tunnel.

governor to be compared with that of the simulation using the governor model and will serve to further validate the model. More importantly however the model based tests using the real governor, will help establish the optimum settings for the governor under a range of operating conditions before these are tested on the power station itself, thus reducing the risks previously associated with such tests.

6. CONCLUSIONS

System managers are placing increasingly stringent demands of the operation of pump storage hydro-power stations when used for frequency regulation and control resulting in a need for a better understanding of their behaviour in relation to a wide range of system conditions. By creating a comprehensive model of the station capable of representing a wide range of operating conditions and which reflects the interaction between the station and the current grid conditions it has proved possible to increase the level of understanding of governor behaviour and of areas for investigation.

The further development of the model into a form where it is capable of being used as part of a 'hardware-in-the-loop' simulation will support the rapidly evaluation of the behaviour of the real governor under a wide range of system conditions and configurations without risk to the station or its associated plant and will support the optimisation of the governor settings.

ACKNOWLEDGEMENTS

The authors would like to thank the Engineering and Physical Sciences Research Council and First Hydro for their support in carrying out this work.

REFERENCES

1. D.Ramey, *Detailed hydrogovernor representation for system stability studies*, IEEE Trans, PAS-89, 1970, pp 106-112
2. Wozniak, *Conduit representation in closed loop simulation of hydro-electric systems*, ASME publication paper No 71-WA/FE-26, 1972, pp 1 - 6
3. Sanathanan, *Accurate low order model for hydraulic turbine penstock*, IEEE Trans. Energy Conversion, EC-2, No 2, 1987, pp 196 - 200
4. Elgerd, *Optimum megawatt-frequency control of multi-area electrical energy systems*, IEEE Trans, PAS-89, 1970, pp 556 - 563

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Mechanical Earthworm to facilitate the access and repair of buried infrastructure.

W. L. Dudeney and M. R. Jackson

Mechatronics Research Group, Department of Mechanical Engineering, Loughborough University,
Loughborough, LE11 3TU, U.K. Tel: +44 (0)1509 223231 Fax: +44 (0)1509 223934
e-mail: w.l.dudeney@lboro.ac.uk m.r.jackson@lboro.ac.uk

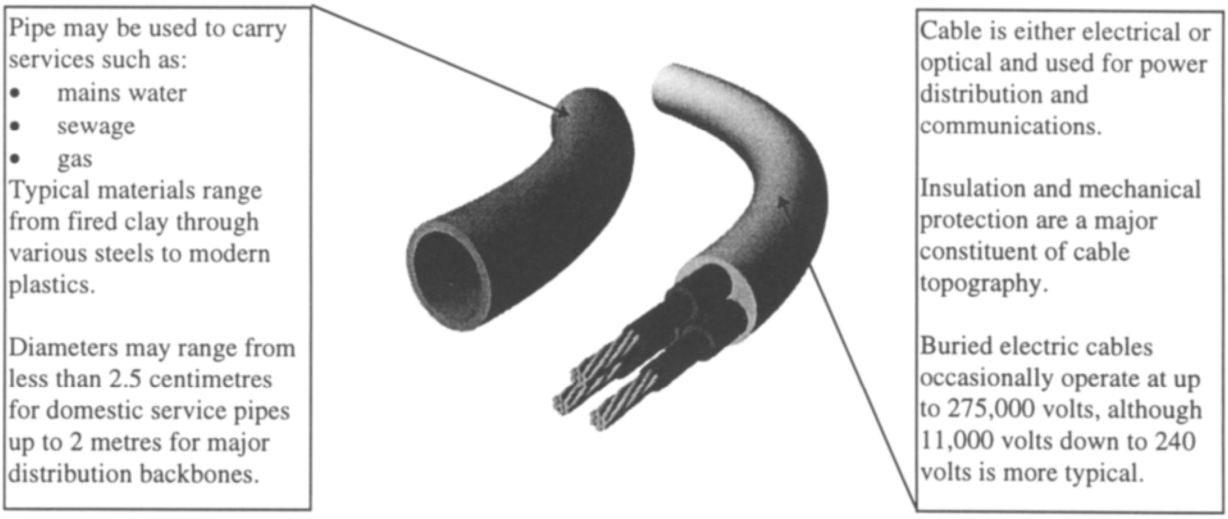
This paper introduces an access problem faced by industries owning buried infrastructure. Current solutions to the problem are assessed and their disadvantages are highlighted. A novel concept solution is proposed which will address current disadvantages, so providing an efficient solution with generic applicability. The proposed solution follows a keyhole surgery principle with an in-hole propulsion mechanism. The system is briefly described with additional detail given to biomechanical analogies used in the propulsion mechanism.

1. INTRODUCTION

There are four main owners of buried infrastructure: gas distributors, electricity distributors, water distributors/sewage collectors, and communications providers. Infrastructure types are pipe and cable (see figure 1). This infrastructure is buried, in most instances for the convenience of the public and to provide a mechanically and thermally stable bedding.

In normal circumstances it is unnecessary for the operator to physically access this type of infrastructure. However various problems do arise for different reasons, often created by third party intrusion such as subsidence caused by heavy road vehicles mounting pavements, and sometimes by operational wear such as gradual degradation caused by encrustation of a cast iron water main. When problems occur it is often necessary to access the infrastructure in question, the challenge is to provide this access whilst minimizing urban disruption.

Figure 1. Two types of buried infrastructure.



Pipe may be used to carry services such as:
- mains water
- sewage
- gas

Typical materials range from fired clay through various steels to modern plastics.

Diameters may range from less than 2.5 centimetres for domestic service pipes up to 2 metres for major distribution backbones.

Cable is either electrical or optical and used for power distribution and communications.

Insulation and mechanical protection are a major constituent of cable topography.

Buried electric cables occasionally operate at up to 275,000 volts, although 11,000 volts down to 240 volts is more typical.

2. STRATEGIES TO ACCESS BURIED INFRASTRUCTURE

Historically, access to any buried object has involved some form of excavation. Modern techniques still rely heavily on excavation of overlying and surrounding material, but in specific circumstances it is possible to gain access non-intrusively if the infrastructure concerned allows internal access (e.g., pipes with sufficient internal diameter).

Current strategies to rehabilitate under-performing buried infrastructure are many and varied, but the more novel technologies can be shown to be [4,5] specific rather than widely applicable. A core decision which will determine the nature of any access required is whether to replace a section or refurbish a section as the processes may be very different. Buried distribution and collection systems often run beneath other urban services such as road ways and pavements, it is for this reason that access is difficult and likely to cause urban disruption.

2.1. Repair of buried cable

Power distribution cables are usually bedded directly in the ground, whilst modern optical communications networks are ducted in 15 centimetre diameter plastic pipe. Cable affords no internal access so excavation is the only method of providing access. Cable faults are failing insulator or failing conductor, or both. There are established techniques [1] to accurately locate points of failure to within 4 metres in typical circumstances. It is normal to replace a 2 metre section of cable around any point of failure. This is skilled work and is currently a job requiring hands-on labour (implying an excavation large enough to allow human access). Excavation of this nature is a common sight in urban U.K.

2.2. Repair of buried pipe

Pipe is bedded directly in the ground. There are many different types of pipe in service, from fired clay, concrete, and cast iron, through to steel and plastic. Typical problems with pipe are blockage, corrosion, and failing joints and valves. There are well developed sonic techniques for leak detection [2,3], but the accuracy of pinpointing leak detection is variable dependent on factors such as: leak velocity, pipe material, mechanical qualities of surrounding ground. Pipe offers the possibility of internal access. There are systems in use to allow inspection, refurbishment and replacement of pipe. Briefly, these include in-pipe robotics for inspection, lining mechanisms, and various systems to replace the pipe with similar, smaller, or even larger diameter pipe. All these systems operate within the pipe, requiring at most an entrance and an exit point. These systems are detailed elsewhere [4,5].

However, Many pipes are too small to allow internal access, and often the nature of the problem may not suit any of the internal repair systems (for example leaking joints and valves). In these instances excavation is required to allow skilled manual labour (implying a man-in-hole sized excavation).

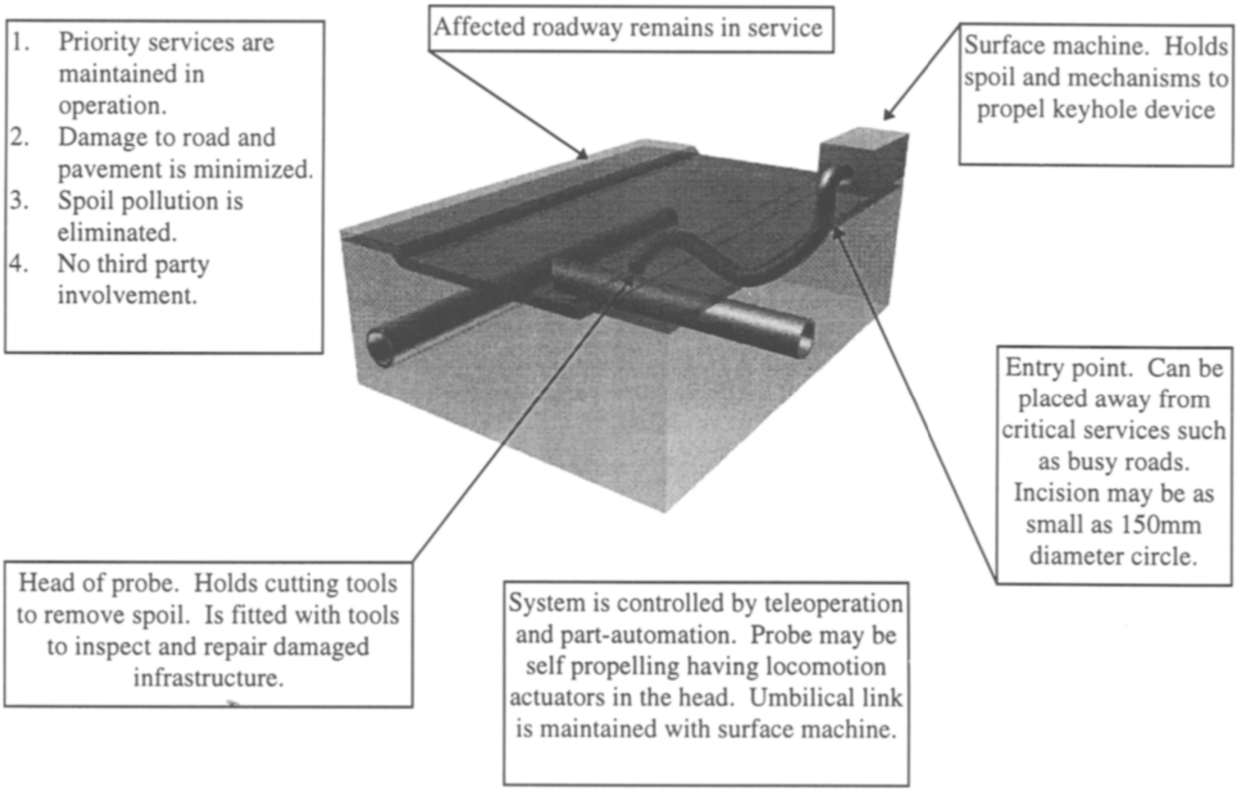
2.3. Evaluation of current strategies

Although no data is available (inquiries were made at Department of Environment, Department of Transport, and many owners of buried infrastructure, as well as their regulators), it is easy to accept that the majority of buried infrastructure failings are accessed through open excavation. There are economic, social, and environmental disadvantages to open excavation in urban areas:

- Urban disruption includes road works generated to allow access to services beneath the road.
- Spoil becomes distributed.
- Excavation produces noise.
- Excavation often leads to permanent and progressive damage to roads and pavements.
- Excavation often requires plant hire.
- Third parties are required to reinstate surface structures.
- Excavation often damages other buried infrastructure.

Despite these disadvantages, excavation continues to be popular. This might be simply for the reason that repair operations are technical and require experience and hands-on skill to perform.

Figure 2. Concept of keyhole surgery applied to buried infrastructure.



1. Priority services are maintained in operation.
2. Damage to road and pavement is minimized.
3. Spoil pollution is eliminated.
4. No third party involvement.

Affected roadway remains in service

Surface machine. Holds spoil and mechanisms to propel keyhole device

Entry point. Can be placed away from critical services such as busy roads. Incision may be as small as 150mm diameter circle.

Head of probe. Holds cutting tools to remove spoil. Is fitted with tools to inspect and repair damaged infrastructure.

System is controlled by teleoperation and part-automation. Probe may be self propelling having locomotion actuators in the head. Umbilical link is maintained with surface machine.

However, with developments in teleoperation [6] it should now be possible to remove the need to physically place the skilled labourer at the work site and so reduce the size of any excavation.

3. THE MINIMAL INTRUSION CONCEPT

The ideal way to minimize the disadvantageous effects of excavation is to adopt a key hole surgery principle. The key hole concept meets the desire to avoid disturbance of critical factors around the work site. It would be possible to access infrastructure below a road way without affecting that road in any way (see figure 2).

The device would be 'inserted' into the ground at a convenient location, away from critical services. From this point the device tunnels its way towards the desired location under the control of a human operator. On arrival at the desired location the damage is inspected and repaired by the operator, using a selection of on-board tools and materials. Following a successful remote repair the device is withdrawn along the tunnel it has created, back-filling as it goes. The only surface evidence of the operation will be a small circle where the device was inserted. At all times waste materials and spoil are managed. Noise will be reduced because cutting operations will be underground.

4. THE MECHANICAL EARTHWORM

The construction of a feasible plan for the key hole surgery principle requires ambitious thinking. The desire is to create a self sufficient excavation and repair system with wide applicability and the ability to address many if not all the disadvantageous factors in traditional excavation.

From this the concept can be detailed as a system including: a tunnelling vehicle, a surface power and spoil/materials handling unit, a human control interface, an umbilical connection between the surface unit and the vehicle, and a set of interchangeable tools for tunnelling and repair/inspection. A graphical representation of this system is given in figure 3.

It is accepted that there are technological challenges in many areas of this system, some of those identified are:

- in-tunnel propulsion,
- underground vision,
- cutting through urban ground,
- directional tunnelling,
- obstacle identification,
- teleoperation procedures/human interfacing,
- interchangeable teleoperation tool design,
- vehicle extraction and ground reinstatement.

Brief consideration has been given to a possible topography for the tunnelling vehicle, this is best described by figure 4. Other than this the current

Figure 3. Schematic representation of the conceptual key-hole surgery system's main components.

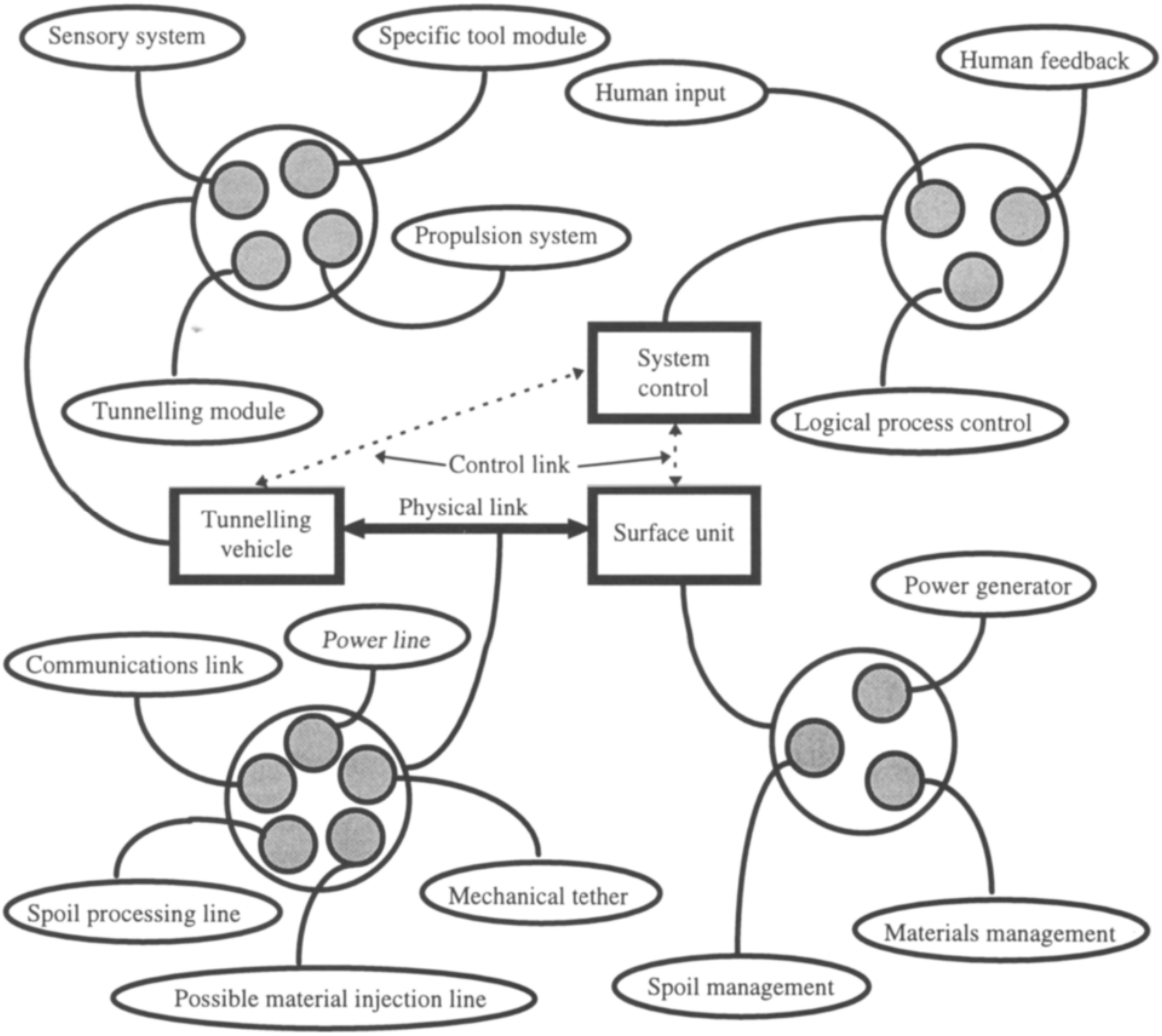
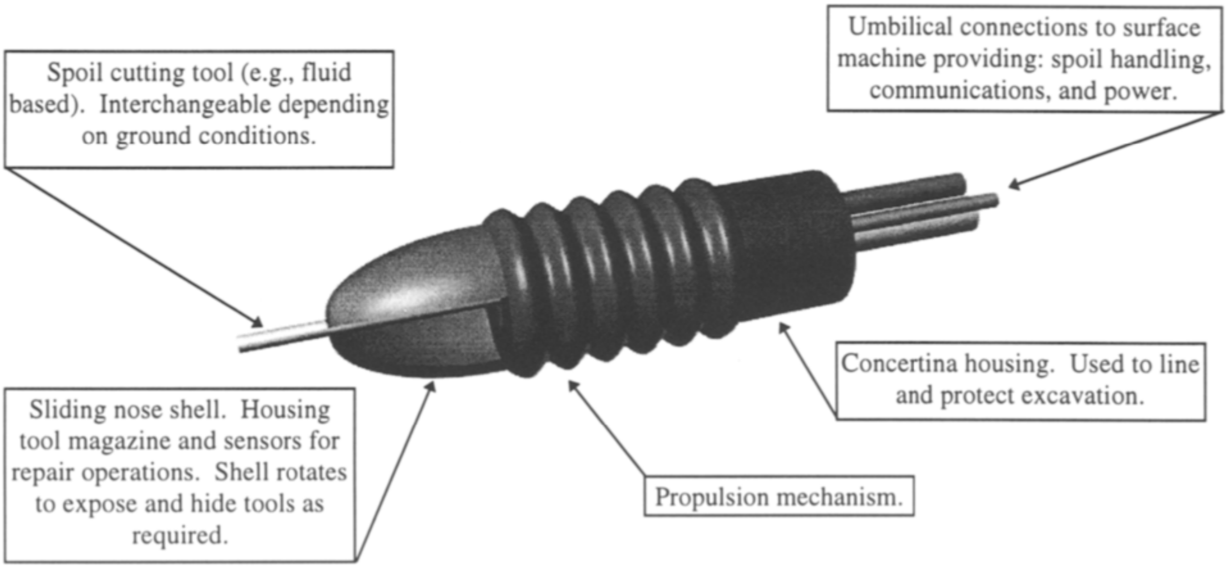

Sensory system
Specific tool module
Human feedback
Human input
Propulsion system
System control
Tunnelling module
Logical process control
Control link
Physical link
Tunnelling vehicle
Surface unit
Power generator
Power line
Communications link
Mechanical tether
Spoil processing line
Materials management
Spoil management
Possible material injection line

Figure 4. Visual of possible topography of tunnelling vehicle.


Spoil cutting tool (e.g., fluid based). Interchangeable depending on ground conditions.
Umbilical connections to surface machine providing: spoil handling, communications, and power.
Concertina housing. Used to line and protect excavation.
Sliding nose shell. Housing tool magazine and sensors for repair operations. Shell rotates to expose and hide tools as required.
Propulsion mechanism.

research focus is on in-tunnel propulsion and the conclusion of this paper will concentrate on a novel method.

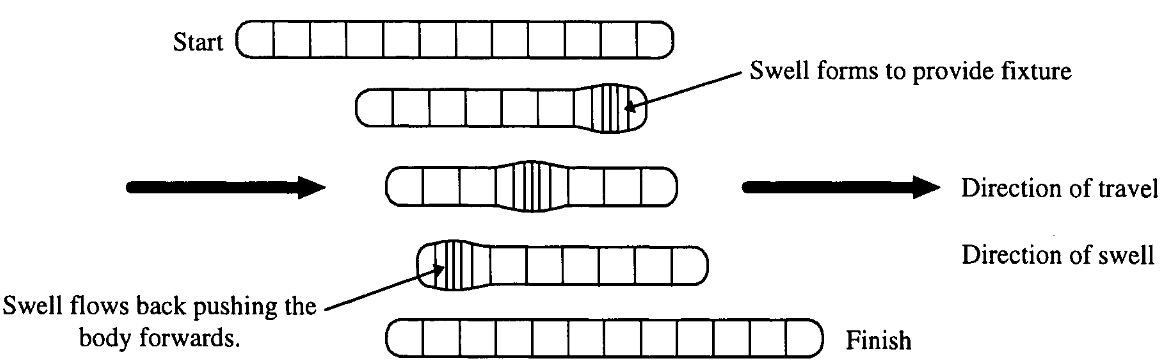
5. BIOLOGICALLY INSPIRED IN-TUNNEL PROPULSION

Biological inspiration has been sought to address the issue of providing propulsion. The locomotion mechanics of the earthworm is a suitable system on which to base an artificial in-tunnel propulsion mechanism. Basic earthworm locomotion is described in figure 5, and is described in greater detail elsewhere [7,8].

The earthworm is analogous to an elastic sack containing a fixed volume of fluid. The elastic sack includes embedded antagonistic muscle groups which operate to pump fluid between different segments of the sack. Because the volume of the sack is constant any swell in one location must be taken up by either shortening of that location, or by constriction of another location. Alternatively the muscle groups may be used to elongate one section and maintain local volume by constricting the section. Articulation may be achieved by antagonistic operation of muscle groups which run the length of the sack (see figure 6).

5.1. Mechanical implementation of earthworm

Figure 5. Description of basic earthworm locomotion.


Start
Swell forms to provide fixture
Direction of travel
Direction of swell
Swell flows back pushing the body forwards.
Finish

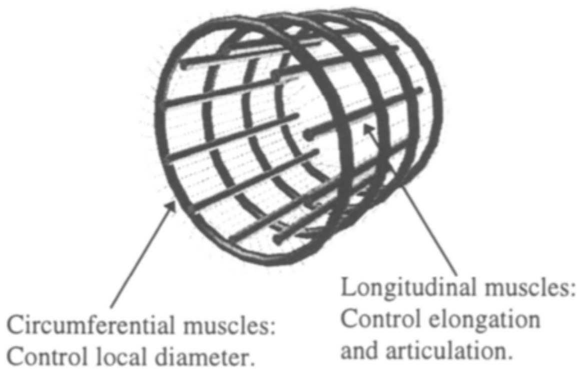


Figure 6. Worm segment sketch to show muscle groups used in locomotion.

locomotion principle

Three priority issues are recognized in the mechanical implementation of an earthworm locomotion system:

- Power supply type
- Actuator requirements
- Closed loop control

Current work is focused on an embedded/surface mounted positional feedback system to allow a control system to know the deformation shape of an elastic surface. Initially it is sought to provide 1 and 2 dimensional information and subsequently through wrapping the surface into a cylindrical form to interpolate 3 dimensional information.

REFERENCES

1. T. Gonen, Electric Power Transmission System Engineering. 1998.
2. E. S. Cole, Methods of Leak Detection: An Overview. Distribution Systems - Actions and Innovations, AWWA, 1980.
3. G. L. Laverty, Leak Detection: Modern Methods, Costs, and Benefits. Distribution Systems - Actions and Innovations, AWWA, 1980.
4. M. Najafi, State-of-the-art Review - Trenchless Pipelind Rehabilitation Systems. 2nd International Conference on Advances in Underground Pipeline Engineering, 1995.
5. T. Takana, The Application of Pipe Repair, Rehabilitation Techniques and their Recent developments. Distribution/Transmission Conference of the Operatins Section of the American Gas Association, 1986.
6. R. Kalawski, The Science of Virtual Reality and Virtual Environments, 1993.
7. D. Nichols, The Oxford Book of Invertebrates, 1971.
8. T. Newton, Muscling in on Nature. EPSRC Newsline, Winter 1997.

Some information presented in this paper has been sourced directly from relevant industries and regulators in the U.K. These organisations include:

Scottish Power, The York Waterworks, Yorkshire Water, Transco, London Electricity, Bristol Water, OFWAT, OFTEL, OFFER, South Western Electricity, Thames Water, Southern Water, London Cable.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Training of experts in Mechatronics in Moscow Bauman Technical University

N.A.Lakota, E.D.Gorbatchevich and V.S.Medvedev

Mechatronics and Robotics Department, Moscow Bauman State Technical University
2 Baumanskaja st. 5, Moscow, 107005, Russia

This paper provided you with the structure of the training courses on trans diciplinary qualification-Mechatronics. Groups (cycles) of training objects: humane-social-economical, mathematical and natural science, mechanical (including electromechanical and hidromechanical), electronics and computer science, automatic control theory (including artificial intelligence), mechatronics systems design, mechatronics systems manufacturing and management and contents of main methods of training have been considered.

1. INTRODUCTION

In connection with development of scientific researchers and new practical development there was the requirement for preparation of the experts, in particular of engineers of Mechatronics. The corresponding courses have been founded in the universities of many countries with this purpose. According to the standard the purpose of training on a "mechatronics" speciality is preparation of the engineers in research, development, manufacture and operation of the controlled system with complex mechanics and advanced control systems, which are constructed on the base of computer and electronic technology, in Russia. Training of experts for the "self-aligning drivers" speciality was started at Moscow State Technical University named after N. E. Bauman in 1951. The training included the courses of mechanics (electro and hydromechanics), electronics and theory of automatic control. Graduate has degree of the engineer-electrical mechanic. Further, the number of hours to study the electronic devices and the systems were increased and new courses of the application of the computers and microprocessors in the modern technical systems and devices were introduced. It was connected with the development of the electronics and the computer technology and the wide implementation of them in the industry. The number of hours to study the theory of automatic control were increased too. As a result, a student studying this course has received necessary volume of knowledge to be an expert in mechatronics in modern understanding of this term. New speciality named "mechatronics" has founded at the university in 1997.

2. GENERAL STRUCTURE OF TRAINING

We shall consider the basic organisation principles of the educational plan in Bauman MSTU. The total educational time of a student at the university is 6 years and 9 months. According to the educational plan the study courses are subdivided into four types: humanitarian and socio economic, mathematical and common natural sciences, general professional and special. The distribution of the number of the hours on kinds of courses is shown at Fig. 1.

Marking and numeration of the courses type at figures 1 and 2 correspond of numeration of the courses at tables 1.

Such distribution of tutorial time is typical for the number of university courses. Common courses are studied during first and second years of training. According to the educational plan the training of educational material differentiated by time and themes is studied on higher courses. The structure of the educational plan for the "mechatronics" course is considered in detail below.

3. TRAINING OF SPECIAL COURSES

Proceeding from the put purpose of preparation of the experts on mechatronics, course of training of students it is possible to group in cycles:

Table 1
Distribution of education time on kinds of courses

№	Kind of Courses	Number of hours			
		4 years of training		6 years of training	
		Total	Lectures	Total	Lectures
1	Humanitarian and socio economic	1548	864	1768	978
2	Mathematical and common natural sciences	1944	1278	2242	1188
3	General professional	1998	1278	3844	2478
4	Special	828	485	2510	1298
	Total	6228	3905	10346	5942

humanitarian and socio economic; mathematical and common natural sciences; theory of automatic control; mechanical electrical-, hydro- and pneumomechanical elements and devices of mechatronical systems; electronic and computing devices of mechatronical devices and systems; technology and organisation of manufacture of mechatronical devices and systems.

The structure of the training process on mechatronics can be divided on three stages.

First stage (first and second courses, 72 weeks of theoretical studying, 12 weeks of examinations) includes studying: of common natural sciences disciplines; introduction of design of the mechatronical devices; base of general technology of the manufacture in machine building and device building fields

Second stage (third and fourth, 69 weeks of theoretical study, 8 weeks of practice, 13 weeks of examinations) includes studying : mathematical base of the control theory; theory bases of the automatic control; main components of mechatronical devices; methods of the design and for manufacture of such systems; and first and second technological practices.

As a result of first training and second training stage students receive necessary knowledge of the basic components of mechatronical systems and the required skills in design of elements of the systems.

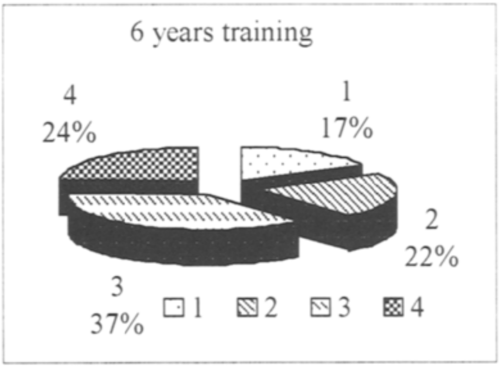


Fig.1 Distribution of education time on kinds of courses

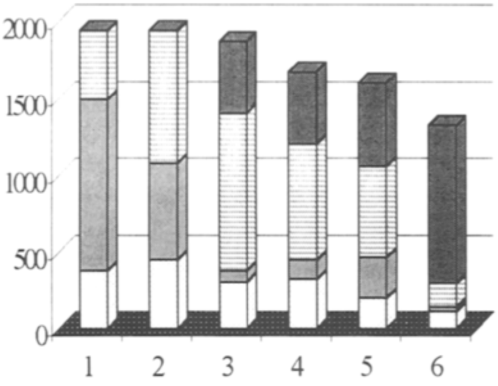


Fig.2 Distribution of education time on kinds of courses

Table 2
Distribution of education time on cycles of courses

№	Cycles of courses	Total number of hours							
		1st year	2nd year	3rd year	4th year	5th year	6th year	4 year	Total
1	Humanitarian and socio economic	378	450	306	324	198	112	1458	1768
2	Mathematical and common natural sciences	846	630	72	126	270	28	1674	1972
3	Mechanical electrical-, hydro- and pneumomechanical devices	0	342	774	54	72	0	1170	1242
4	Electronic and computing devices, programming	270	0	234	450	216	84	954	1254
5	Theory of automatic control	0	72	378	198	108	266	648	1022
6	Design of mechatronical systems	324	90	108	234	414	768	756	1938
7	Technology and organisation of manufacture of mechatronical systems	126	360	0	288	324	70	774	1168

A student gets the bachelor's degree after finishing the fourth course and defence of the qualifying project.

The third stage (5-6 courses, 50 weeks of theoretical study, 9 weeks of practice, 8 weeks of examinations, 20 weeks of diploma design) is final stage of training. The modern theory of automatic control (including logical control in mechatronical systems and elements of artificial intelligence), methods of such systems design, microprocessors and programming of them, technology and organisation of manufacture of mechatronical systems are studied during this stage. Three manufacture practices (design, operational and before diploma) are conducted during one.

During all three grade levels the students study the number of socio economic courses to receive knowledge in understanding of a role of engineering activity in the decision of social, economic and political problems and skill of foreign language. After finishing study, state examinations and defence the diploma project the students receive degree engineer (Master of Science) for mechatronics and the diploma of standard form.

Table 3
Distribution of educational time on cycles of courses and methods of training

Courses	Total	Lectures	Laboratory works	Practical works	Independent works
1. Humanitarian and socio economic	1768	348	216	414	790
2. Mathematical and common natural sciences	1812	720	108	414	576
3. Theory of automatic control	1022	404	86	150	382
4. Mechanical electrical-, hydro- and pneumomechanical devices of mechatronical systems	1080	396	126	90	468
5. Electronic and microprocessor devices of mechatronical systems, programming	1254	492	138	90	534
6. Design of mechatronical systems	2330	572	124	144	1490
7. Technology and organisation of manufacture of mechatronical systems	1132	456	234	104	338
Total	10404	3290	1106	1342	4578

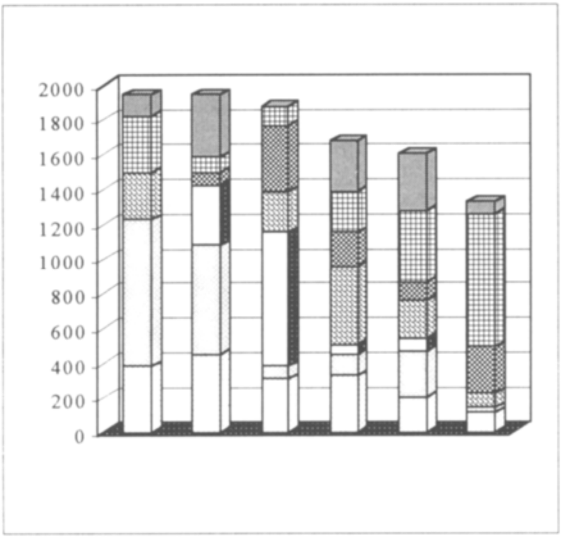


Figure 3 Distribution of educational time on cycles of courses

4. METHODS OF STUDY

Distribution of the number of hours under methods of study (lectures, exercises, laboratory works and independent study of material) is shown at table 3 and at Fig 4 and Fig. 5.

Number of sections at Fig.4 corresponds the following methods: 1- lectures, 2- exercises, 3- laboratory works, 4- independent study of material. Numbers shown at Fig. 5 describe the cycles of courses in according with them numeration at the table 3. Order of rows in each cycle (from left to right) is analogous to numeration of the study methods.

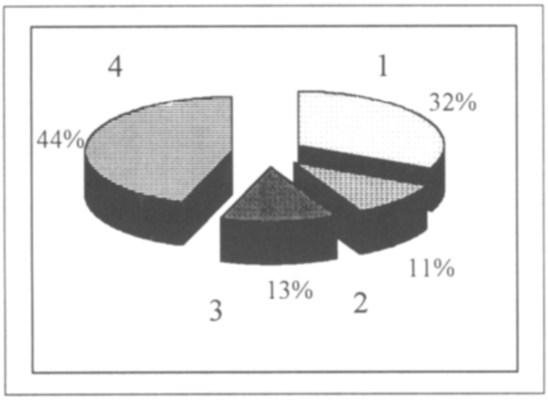


Fig. 4 Distribution of total time for the study methods

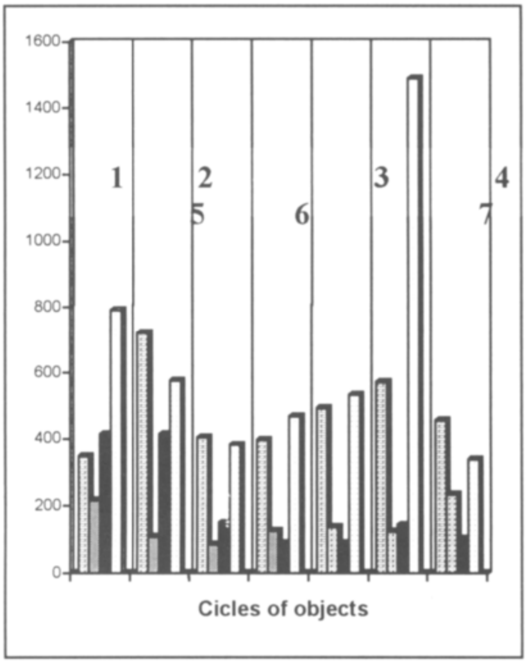


Fig. 5. Distribution of total time for cycles of courses and for the study methods

Main study methods of theoretical base of speciality are lectures and seminars, which spend 43 % of total educational time. Great attention is devoted to independent work of the students (44% of total educational time). During independent work the students works under materials of the lectures, make home works and carry out course and diploma projects.

During the study a student makes 4 course projects and one course work. During study on the third course a student designs the mechanical mechatronical device (as a rule it is reducer for connection of conducting motor with object or sensor). During study on the fourth course a student makes the work on calculation and realisation of correcting devices in actuators of displacement of conducting mechanism. The other project is design of the electronic amplifier. During study on the fifth course a student defences the project for design of manufacture process for the creating of mechatronical unit (mechanical or electronic) and the design project of documentation creating for the individual mechatronical device (self-aligning driver, manipulator of mobile robot with control system and so on). During study on sixth course a student produces the programming of the microprocessor project, which controls the mechatronical systems or part of it. Diploma project consists of the theoretical, design and technological parts. It includes of calculations, design and manufacture of the units of the mechatronical systems. Ecological and economical aspects must be considered at the project too. The typical themes of the diploma projects are "automatic pilot of aeroplane", "driver of guiding for the radio locator antenna", "control system of piezo electrical motor", "control system of manipulator for the mobile robot", "control system of the automatised bakery".

During study great attention is devoted to receiving experience in regulation and research of mechatronical systems by students. This problem is solved by carrying out of the laboratory works and industrial practices. Besides usual laboratory works for courses of the physics, chemistry details of machines, strength of materials, electrical engineering and so on. Study for research of the mechatronical systems (such act as automatic pilot of aeroplane, driver of guiding for the radio locator antenna, control system of the mobile robot",) are carried out too. For example, the unit for research of the control systems of the conducting electrical motors of different types (direct current, double and three phase synchronous , step by step electric and piezoelectric motors) is shown at Fig 6. The unit is the universal platform. Motors of different types, sensors (velocity and displacement), reducers with imitation of dynamic and static load and moment loader for imitation of external moment, which can be changed arbitrary, can be fixed at this unit. Motor connects with the computer via electronic units. Control of motor (and another elements of the unit) and results proceeding are realised via this computer.

Some laboratory works are carried out with use this unit. Students solve the following tasks - definition of the parameters for the motors and sensors, determination of the characteristics of the electronic control schemes of motors, research of drivers with feedback with different control, programming of the control by means of computer during these laboratory works. During study the students take part in three manufacture practices: during third course (4 weeks - manufacture in machine building), during fourth course (8 weeks - design and manufacture of mechatronical devices and systems), during fifth course (5 weeks exploitation of mechatronical devices and systems).

During sixth course diploma practice is carried out The aim of it is to find materials for the diploma project. The test of students knowledge is realised by means of the defence home works, course projects and works, by examinations and tests. during study a student has 59 tests and 44 examinations.

Experience shows that specialists of mechatronics are required in different areas of industry and graduates find work very easy.

REFERENCES

1. Milne J.S., Frazer C.J. Development of Mechatronics Learning Facility. Mechatronic Systems Engeneering, **1**, 1990

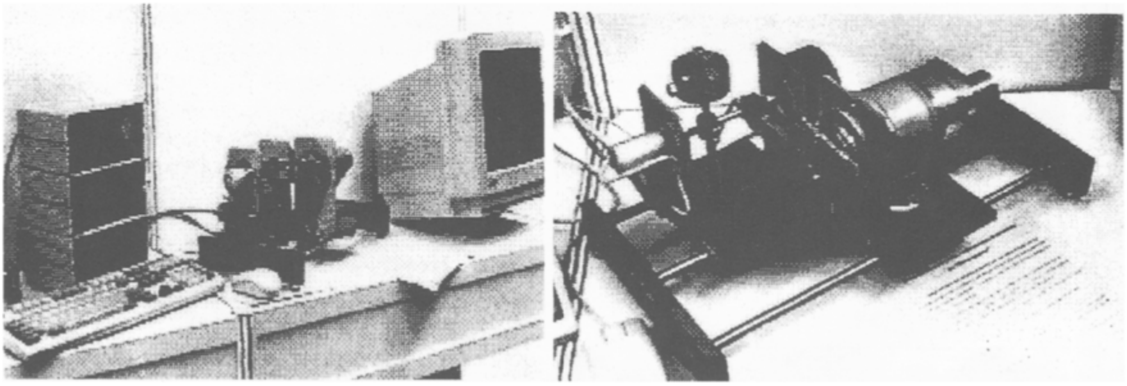


Fig.6 Laboratory unit for research of electric motors into the system computer-electronic interface- electric motor.

2. Yamazaki K., Miyazawa S. A Development of Coursware for Mechatronic Education. Int.Journ. Appl. Engineering Education. Vol.8, No 1, 1992, pp. 61-70.
3. Milne J.S., Frazer C.J., Logan G.M. Mechatronics in Engineering Courses. International Journal of Mechanical Engineering Education, vol 20, No. 2, 1993.
4. Fraser C.G., Milne J.S., Logan G.M. An Education Perspective on Applied Mechatronics. Mechatronics, Vol. 3, No 1, 1993, pp. 49-57.
5. Csibi V., Maties V. Mechatronics and Engineering Education at the Technical University of CLUJ. Proc. Joint Hungarian British Int. Mechatronics Conf. Budapest, 1994, pp. 771-774.
6. Vodovozov V.M. The Educational Resources of Mechatronics. Mechatronics Vol. 5, No 1, 1995, pp. 15-24.
7. Pavlov V.I., Avramov I.I., Trojanov B.P. Specializing Training on Mechatronics in Field of Mechanical Engineering. Proc. Int. Conf. on Resent Advances in Mechatronics, Istanbul, Turkey, 1995, pp. 283-287.
8. Acar M. Mechatronics Engineering Education and Training. Proc. Int. Conf. on Resent Advances in Mechatronics, Istanbul, Turkey,August 1995, pp. 238-239.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronics as an engineering science

Jan Wikander and Martin Törngren

Mechatronics Lab, Department of Machine Design
Royal Institute of Technology, S-100 44 Stockholm, SWEDEN
e-mail: jan/martin@damek.kth.se

Abstract. The paper discusses the current understanding of mechatronics as an interdisciplinary subject. A conclusion is that mechatronics tends to attract contributions from many related fields, leading to a lack of focus as exemplified by many mechatronic conferences. Instead, the authors claim that engineers and researchers in mechatronics should concentrate on the opportunities and challenges arising specifically due to the interdisciplinary interactions which are characterizing mechatronics. Current practice in mechatronics engineering is characterised by a subsystem based approach by which integrated systems are built from technology homogeneous subsystems without a real demand on development of a certain technology as a result of its closer integration with other technologies. As opposed this it is claimed that *mechatronics as an engineering science* should focus on the interdisciplinary interactions, and based on these identify, formulate and conduct new research. The paper presents a set of new research directions, research results and grand challenges relating directly to interdisciplinary interactions between mechanical engineering, control engineering and computer science.

1. INTRODUCTION

So far there is no common and widely accepted understanding of what mechatronics really is. Many different notions similar to or including mechatronics have been used in various contexts; micro-mechatronics, opto-mechatronics, super-mechatronics, mecano-informatics, contro-mechanics and megatronics are some of those, each coined to put forward a specific aspect or application of mechatronics. Examples of attempts to describe mechatronics include:

- Mechatronics encompasses the knowledge and the technologies required for the flexible generation of controlled motions [1].
- Mechatronics is the synergistic combination of mechanical and electrical engineering, computer science, and information technology, which includes control systems as well as numerical methods used to design products with built-in intelligence [2].
- Hewit in [3] importantly states: A precise definition of mechatronics is not possible, nor is it particularly desirable, because the field is new and expanding rapidly; too rigid a definition would be constraining and limiting and that is precisely what is not wanted at present.

Mechatronics as an interdisciplinary subject tends to attract contributions from all related fields without really putting forward the opportunities and challenges arising specifically due to the interdisciplinary interactions. An example of this is that many mechatronics conferences have been un-focused and thereby not attracting the most adequate contributions, which definitely exist. This is a disadvantage in that it hampers the development of mechatronics as an engineering science. Scientific publications in mechatronics, to help in making the subject more focused, are still quite rare. One of the earlier is Mechatronics - an International Journal published by Elsevier Science first published 1991. A more recent publication, the IEEE/ASME Transactions on Mechatronics appeared in 1996.

This contribution is not just another attempt to describe what the research community means with the term mechatronics. Rather, we try to get to the heart of multidisciplinary engineering of which mechatronics is an excellent example, and point out how the integration of disciplines leads to new degrees of freedom in design and corresponding research directions which otherwise would not have been investigated. This is the major contribution achieved by a multi-disciplinary approach to engineering science; it leads to a new important research field and at the same time it helps to push research in related fields into new fruitful directions.

A general discussion on interdisciplinarity in research, its claimed lack in academia and its tremendous importance for the next century is discussed in detail by Mayer-Krahmer in [4]. Large studies are referred to in which it is concluded that the technology at the beginning of the next century cannot be separated according to conventional disciplines and further that important innovations often stem from the interaction of several previously unconnected streams of scientific and technological activity. We claim also that the full potential of interdisciplinarity also includes bridging the gap between real applications including experimental research and the scientific disciplines. Also this is discussed in [4] and referred to as the dynamics of research and technology; a first science-push cycle is followed by a demand-pull cycle, a process being non-linear and requiring efficient feedback and interaction.

2. THE PREVAILING VIEW ON THE ENGINEERING OF MECHATRONIC SYSTEMS

Engineering of mechatronic systems and products is well established in a substantial number of industrial branches like automotive, manufacturing systems, aircraft control, construction equipment etc. Such engineering typically apply a subsystem based approach to mechatronics. By *subsystem based* we refer to a product development strategy by which integrated systems are built from technology homogeneous subsystems (mechanics, electronics, control and software). The subsystems are developed in a concurrent manner with an important focus on subsystem interfaces. Once the interfaces are designed, each subsystem is designed in a fairly traditional way. This means that the focus has been on team building to improve communication and multidisciplinary understanding between engineers of different expertise such that the interfaces can be properly defined.

In the subsystem based approach there is no real demand on development of a certain technology as a result of its closer integration with other technologies, e.g. the close integration of automatic control and computer science. The performance of the mechatronic system is instead merely a result of a sound integration of existing technology. In the existing engineering literature on mechatronics the subsystem based approach is likewise totally predominant, but with a too limited coverage of the development process and corresponding team building. Typically, books of this literature devote the first chapter to defining or explaining what mechatronics is, and than the remaining chapters cover a subject each (modelling, sensors, actuators, control, computer hardware, interfaces, communication etc.) in a traditional but short form way, see e.g [2, 5]. Successful mechatronics engineering can hardly be based solely on such literature.

The component based approach to mechatronics is still a drastic improvement from the early days when the mechanical engineers first designed the mechanical system, which was then handed over to the control engineers doing a control design. Concurrently a computer system was designed by the electrical engineers and finally programmers were given the impossible task of designing and implementing a complex controller due to an odd mechanical design, on a too slow computer system.

3. OUR VIEW ON THE ENGINEERING OF MECHATRONIC SYSTEMS

The advances in digital electronics have enabled the possibility to invent, create or improve systems which rely on mechanical components to perform their intended *dynamic* actions. The key disciplines to be mastered concurrently and in an integrated manner are mechanical engineering, software engineering, control engineering and computer engineering. The major shift of paradigm enabled by mechatronics is that of shifting the implementation of functionality from mechanical hardware to computer software, but still and most importantly the end effecting components are mechanical. Notably, we consider software, rather than microelectronics or microprocessors, as being the major new paradigm as it is the software that provides the new and extensive flexibility and freedom in design. However, in many cases the actual software design is implemented in electronic hardware (hardware/software co-design) but in both cases there is a software design level.

Some figures from a world leader in industrial robot design and manufacturing show clearly the trend of moving functionality from mechanics to software. About ten years ago the development engineers were split between mechanics, electronics and software in the roughly estimated shares 60, 25 and 15 respectively, whereas today (1998) the shares are 30, 15 and 55 respectively. Similar figures have been presented for many embedded control system applications related to mechatronics.

A fundamental aspect of mechatronics is that theories and concepts for mechanical design that have evolved over centuries are about to be replaced by software solutions implemented in embedded microcomputer systems with only a fraction of maturity. This is an aspect of significant importance in many applications, especially those that are safety critical. Related to this, the key feature of software is its versatility and flexibility which, unfortunately, easily leads to complexity problems requiring special attention.

The relative youth of software and computer engineering is very apparent by the following authentic example, again from a manufacturer of industrial robots. Back in the late 60's and early 70's a thorough pre-study was performed to investigate the possibilities of designing an industrial robot. The project team finalized the pre-study and presented two competing solutions with respect to implementation of the control system. The two solutions were; an analogue machine and a micro processor system. Note that this was simultaneous with the first microprocessors reaching the market. The conclusion of the pre-study was that the two solutions were technically and economically feasible and equal but that the project team felt more for the micro processor solution. Fortunately, the company made the right design decision and is currently one of the world-leading robot manufacturers.

Similar examples exist, but with less fortunate outcome. A famous example of the same time is a former (!) manufacturer of mechanical calculating machines. The company was faced with the new technology of microelectronics, and made a legendary management pronouncement that they would stick to the mechanical calculators because they believed that the development of the mechanics would be as rapid as the development of the electronics.

The new flexibility in mechatronics design is based upon a number of design choices with respect to how, i.e. with which technology and with which partitioning, certain functionality is implemented. This choice of how functionality is implemented determines the characteristics of transfer functions, data flows and energy transfer. As the choices regarding partitioning of functionality and corresponding implementation technology are made early in the design chain, there are limited possibilities to take resulting implementation characteristics into account to further improve the design. In order to improve this situation there is an urgent need for theories, models and tools which facilitate modelling, analysis, synthesis, simulation and prototyping of multi-technological systems of which mechatronic systems is an important example. Ideally this will lead to a development strategy involving more of conceptual and top-down design and less of ad hoc and bottom-up design.

4. THE ROLE OF MECHATRONICS AS AN ENGINEERING SCIENCE

As opposed to the subsystem based approach.to mechatronics we claim that *mechatronics as an engineering science* should focus on the interdisciplinary interactions, and based on these identify, formulate and conduct new research. Consequently the important role of mechatronics is to bridge the gaps between related engineering sciences. In doing so, limitations due to these gaps become apparent. This has the effect that the application of a mechatronics and thus multidisciplinary perspective will assist in the identification of new research directions in individual disciplines. A fruitful interpretation of this is that mechatronics engineering is about mastering a multitude of disciplines, technologies and their integration/interaction, whereas mechatronic science is about invention and development of new theories, models, concepts and tools in response to needs evolving from interacting scientific disciplines. The corresponding scientific scenario is illustrated in figure 1.

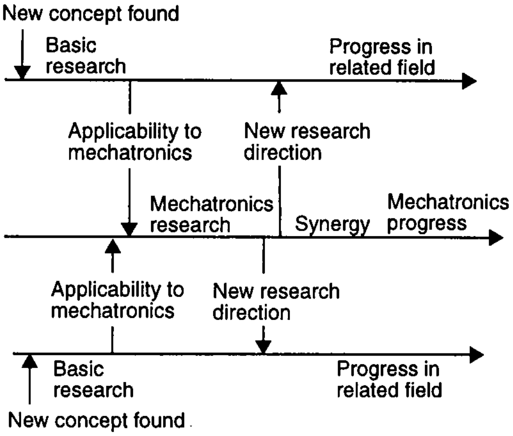

New concept found
Basic research
Progress in related field
Applicability to mechatronics
New research direction
Mechatronics research
Synergy
Mechatronics progress
Applicability to mechatronics
New research direction
Basic research
Progress in related field
New concept found

Figure 1. Integration of disciplines provides synergi and new research directions.

In the following there is discussion of a set of interdisciplinary research directions which have evolved from a multidisciplinary mechatronics approach and for which new research results are available, but also for which further research is encouraged. The below figure visualizes where the challenges and opportunities arise due to the integration between disciplines and technologies.

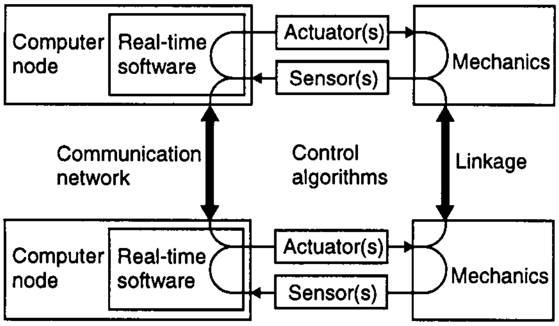

Computer node
Real-time software
Actuator(s)
Sensor(s)
Mechanics
Communication network
Control algorithms
Linkage
Computer node
Real-time software
Actuator(s)
Sensor(s)
Mechanics

Figure 2. Interactions within an integrated mechatronic system.

It is clear from figure 2 that the control algorithms and thus control engineering is essential to mechatronic systems. As has been discussed in the introductory part the shift of paradigm is that of moving the implementation of functionality from hardware to software. A typical design scenario is that an early conceptual design phase results in a course partitioning of functionality between mechanics and control software. Based on this, an initial mechanical design is performed. This design provides a basis for modelling of system dynamics which in turn provides input to the control design. The control design is to be transformed into a real-time software design to be implemented on a computer system which typically is distributed.

The strong integration of technologies (mechanics, control, software, electronics) leads to interdependencies which require special attention in terms of development strategy. A pure sequential development strategy is clearly not suitable, neither is a traditional concurrent engineering approach. The reason is that the strong interdependencies may have substantial influence on for example performance and reliability in a way which can not be forecasted without a new iterative and highly integrated development strategy. Such a strategy needs to be supported by theories, methods, models and tools which are yet to be developed. To support this need, new research directions are identified below, together with some recent research results.

4.1 Control of nonlinear mechanical systems

In the process of transferring the functionality from mechanics to control, the idea is typically to simplify the mechanical design and/or improve the control performance. This leads to the need for new control methods requiring also better models of the mechanical system. A typical example of this is the control of mechanical systems exhibiting nonlinear friction. The ideal mechatronic solution would be that a deficiency of the mechanical system could be cost effectively compensated by a suitable control engineering solution. Due to the interaction of control and mechanical engineering improved friction models have evolved [6] and model based control methods [7] have been developed to compensate for friction. An application example employing both the recent advances in friction modelling and in control of nonlinear system is the servo control of pneumatic cylinders described in [8]. There is a further need for research on control methods (potentially requiring also new models of nonlinear phenomena) for both low complexity systems like single actuators and complex multi-degree of freedom systems.

A related ***research challenge*** here is the development of control methods and control architectures for extremely complex mechanical systems. Such methods and architectures should also support modularity, scaling and reconfiguration. An application example requiring this is the control of legged locomotion systems for difficult terrain.

4.2 Real-time control systems modelling

To design and implement mechatronic systems there is a need for models that support the description and analysis of systems that incorporate among other things control algorithms, software, hardware, interfaces and mechanics. Co-simulation and links between computer aided engineering (CAE) tools is a step in this direction; such tools however do not support distributed computer system implementation. Most theoretical work on the other hand is highly discipline oriented and has not yet been exploited.

Some exceptions in this regard include the Development framework and the Grape project, see [9] for an overview.

Based on this need the AIDA project [9] was started at the Royal Institute of Technology. The intent is to develop a modelling framework that can describe a control design, the computer hardware and software, and parts of the mechanical structure. To focus the work it was decided that the models should primarily support the analysis of the timing behaviour of control applications when implemented in distributed computer system. The novelty of the work is the multidisciplinary perspective on what the models need to contain, resulting in a combination of existing techniques but also in research to develop modelling features that are missing.

The models, which are intended to be used to complement commercial CAE tools, for each corresponding system describe (control functions, software, etc.) the structure, the internal timing and triggering, and the structural building blocks.

The models in particular allow discipline specific models to be used while links between these different views are maintained. One example of this is a control system block diagram with its required timing and triggering behaviour vs. the computer system implementation in terms of processes and scheduling algorithms for processors and communication resources. Another important point is that the internal behaviour models exist in two versions; for example, a specification timing and triggering diagram specifies the required timing behaviour e.g. in terms of period sampling with tolerances; whilst the corresponding implemented timing behaviour contains actual values as derived from analysis or experiments.

More work remains to verify the models (so far used in two case studies), where the next step is to develop a prototype.

A related ***research challenge*** concerns how to best combine different modelling concepts in order to support the development of complex industrial systems characterized by:

- A combination of event- and time-triggered parts both with respect to requirements and implementation.
- A combination of discrete-event and discrete-time dynamic systems.
- requirements on functionality, timing, reliability, safety, etc. i.e. requiring a number of related specifications and attributes to be included in the models.

4.3 Time-varying control systems

In the field of automatic control, implementation aspects directly related to computer systems have received little attention. When a control system is implemented in a distributed computer system it will be subject, to a lesser or greater extent, to [10]

- time-varying feedback delays and sampling periods,
- transient and permanent errors, e.g. lost data in communications.

From a computer engineering and science point of view, on the other hand, there is very little work that directly supports the modelling and analysis of sampled data systems. By studying the implementation of control systems and related problems from an automatic control and computer engineering point of view it becomes apparent that there is much room for new research, for example dealing with:

- The development of scheduling approaches that directly support the relative timing requirements posed by sampled data control systems, [11, 12] One may note that many useful results exist in the field of real-time scheduling theory but that their application is difficult and ad hoc because of a mismatch of the underlying timing models and requirements.
- The development of methods for design of control algorithms that can manage time-varying

feedback delays and sampling periods (possibly including asynchronous components), [13].

- The development of methods for the derivation of timing tolerances of e.g. feedback delays with respect to performance and dependability requirements [11, 13].
- A theory for systematic handling at the application level of errors due to timing and lost data in sampled data systems. This could be integrated with techniques for error detection and handling used in control engineering.

The expected results would provide designers with new alternatives and thereby the possibility to choose from or to combine solutions from the computer and/or control side. A related ***research challenge*** concerns how and where to best introduce redundancy and diversity (system, software, vs. hardware) in order to build robust and dependable systems.

5. CONCLUSIONS

We have pointed to a number of areas which have benefited from the interdisciplinary perspective and a focus on interactions between disciplines. We believe that there are many other such areas including e.g. dependability, architectures and development tools. Within these areas there exist to some extent discipline oriented methods and techniques, but again by studying the interactions between the disciplines the overlap and lack of capabilities of existing theories and methods can be identified. It is the role of mechatronics as an engineering science, to pursue such integrated research activities to extend the current state of the art regarding the design of complex mechatronic systems.

REFERENCES

1. van Brussel, H. The mechatronics approach to motion control. In proc. of the Int. Conf. on Motion control - The mechatronics approach, Oct. 1989, Atwerp Belgium
2. Shetty, D., Kolk R. A., Mechatronics system design. PWS Publishing Company, 1997, ISBN 0-534-95285-2.
3. Hewitt, J., R. Mechatronics - the contributions of advanced control. In proc. of 2nd Conference on Mechatronics and Robotics. Duisburg/Moers, Germany, September 27-29, 1993.
4. Meyer-Kramer F. (1997). Science-based technologies and interdisciplinarity: Challenges for firms and policy. In Edquist C. (editor): Systems of Innovation. Technologies, Institutions and Organizations, London 1997, pp. 298-317.
5. Bolton, W. Mechatronics - Electronic Control Systems in Mechanical Engineering. Longman Scientific & Technical, 1995, ISBN 0-582-25624-8.
6. Eriksson, B. Optimal Force Control to Improve Hydraulic Drives. Licentiate Thesis, DAMEK research group, Dep. of Machine Design, The Royal Institute of Technology, Stockholm, Sweden. TRITA-MMK 1996:3, ISSN 1400-1179, ISRN KTH/MMK/R--96/3-SE.
7. Canudas de Wit, C., Lischinsky, P. Adaptive friction compensation with dynamic friction model, Proceedings of the 13th IFAC Triennial World Congress, San Francisco, USA, 1996.
8. Xiang, F., J. Wikander, Nonlinear modelling and identification of pneumatic servo, Proceedings of the 17th IASTED International Conference on Modelling, Identification and Control, Grindelwald, Switzerland, 1998
9. Redell, O. Modeling of distributed real-tim control systems. Licentiate thesis, Department of Machine Design, KTH, TRITA-MMK 1998:9, ISSN1400-1179, ISRN KTH/MMK--98/9--SE
10. Wittenmark B., Nilsson J., Törngren M. Timing Problems in Real-time Control Systems: Problem Formulation. Proc. of the American Control Conference, June 1995, Seattle, Washington.
11. Törngren (1995), Modeling and design of distributed real-time control applications. Doctoral thesis, Department of Machine Design, KTH, TRITA-MMK 1995:7, ISSN1400-1179, ISRN KTH/MMK--95/7--SE
12. Cheng and Agrawala (1994), Scheduling of Periodic Tasks with Relative Timing Constraints, Technical Report CS-TR-3392, Dept. of Computer Science, University of Maryland, College Park, Dec. 1994.
13. Nilsson, J. Real-Time Control Systems with Delays. Doctoral Thesis, Dept of Automatic Control, Lund Institute of Technology, 1998, ISSN 0280-5316, ISRN LUTFD2/TFRT--1049--SE.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

The Pacificar Project as an integral mechatronics course component: A concept for linking educational and research activities

Charles Brom

Department of Informatics, Communication and Electrical Engineering, Zürcher Hochschule Winterthur (University of applied Sciences), P.O. Box 805, CH-8401 Winterthur, Switzerland, (bm@twi.ch)

The Pacificar project constitutes an integral component of the «Mechatronic Systems» course at the Zürcher Hochschule Winterthur (ZHW). This paper describes the context in which the Mechatroncis Systems course is positioned, as well as elements of the course and the project itself. A mutual benefit in achieving the goals in lecturing, technology exchange with industry, as well as in applied research and development is produced by the close compounding of lecturing and correlated project types, i.e. educational projects, technology transfer projects and research projects. This is illustrated by the Pacificar project. The main aim of the project is to provide a platform suitable not only for research and development projects, but also for experimental systems which are of practical use in the study of mechatronic systems. Autonomous vehicles are especially appropriate for this purpose. They cover an area of research in which, by means of sensors, actuators and real time data processing, innovative systems capable of finding solutions for difficult tasks can be developed. The know-how gained from solving problems within this field can be directly applied to many other mechatronic systems. The Pacificar project was conceived in collaboration with the Department of Computer Science and Engineering of the University of California, San Diego.

1 INTRODUCTION

We understand Mechatronics as an interdisciplinary field of compounded Engineering sciences. Mechatronic systems acquire signals from their environment, process them and generate signals that control actuators which produce mechanical output, i.e. force, momentum and velocity. In addition, we expect a mechatronic system to have in built a certain amount of intelligence. Autonomous vehicles, sophisticated production machines and the active control of the drive dynamics of automobiles are examples of such systems.

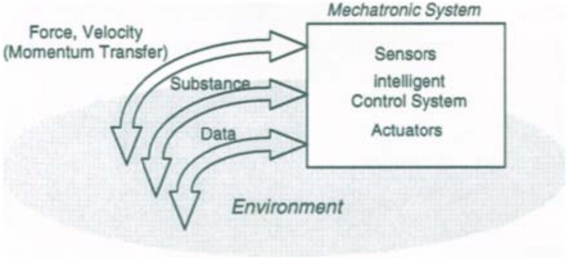


Figure 1. Basic structure of mechatronic systems

The main aim of the Pacificar Project is to provide a platform suitable not only for research and development projects, but also for experimental systems which are of practical use in the study of mechatronic systems. Autonomous vehicles are especially appropriate for this purpose. They cover an area of research in which, by means of sensors, actuators and real time data processing, innovative systems capable of finding solutions for difficult tasks can be developed. The know-how gained from solving problems within this field can be directly applied to many other mechatronic systems.

2 CONTEXT

The Swiss Federal Government has stipulated three main obligations to the Universities of applied Sciences in Switzerland. These are:

- Provision of undergraduate courses for students.
- Provision of postgraduate education, especially relative to continued acquisition of up to date knowledge.

- Provision of knowledge and technology exchange with industry. This also refers to applied research and development (R&D).

Besides the provision of undergraduate courses, technology transfer is a primary obligation for the Universities for Applied Sciences. Technology transfer projects are normally funded by the companies requiring it. Such companies are often of a small or medium size. Funding for research and development (R&D) activities is normally more difficult to obtain. This holds especially true for the initial phase of project ideas, when their significance has not yet been proven. R&D activities, however, are of importance for gaining expertise in new technologies and developments. This expertise is essential for building technology transfer competence and maintaining it at a high level. This paper describes the concepts we are seeking in order to meet the three obligations mentioned above. The main goals are:

- Establishment of a close relationship, and coordinaton between R&D activities and technology transfer projects.
  Efficient use of human resources, infrastructure and funding.
- Integration of undergraduate students in research and development programs with respect to the requirements of education.
- Linkage of research assistants with post graduate studies.

3 CURRICULA FINAL YEAR STUDIES

In the final study year the undergraduate electrical and mechanical engineering students at the Zürcher Hochschule Winterthur (ZHW) establish their own area of study, according to their own needs and interests. They choose two subjects for in-depth coverage out of a menu of subjects. The subjects are lectured over a time period of one year in blocks of four lessons per week. Additionally they are reinforced by four laboratory instruction lessons per week over the first semester of the final year. Examples of subjects with in-depth coverage are, among others: Power Electronics and Drives, Microcomputer Systems, Software Engineering, Control Systems, Communication Networks, Microelectronics.

In addition to the course subjects with in-depth coverage, the students select a minimum of two complementary subjects, from a choice of over twenty; Students who wish to, can also opt for a third complementary subject. The course «Mechatronic Systems», in which the Pacificar project constitutes an integral part, forms part of these complementary subjects.

The subjects with in-depth coverage, as well as the complementary subjects are designed as modular units. Admission is thus open to all undergraduate and graduate students that have studied the basic engineering disciplines. Therefore these courses are also attractive for practising engineers seeking additional education in their particular field.

With the focus on reinforcing and extending their theoretical knowledge, students carry our two individually assigned projects in their chosen field of interest. At the end of the course, after having taken all their exams, the students have eight weeks in which to complete their thesis. Often the theses is continuation of an existing design project begun at an earlier stage of study.

4 GOALS OF THE MECHATRONICS SYSTEMS COURSE

Mechatronic systems cover a wide field and therefore the areas treated in the courses have to be limited. The course teaches those essentials vital to the understanding and designing of mechatronic systems. It will introduce the principles of how of components and subsystems function, and illustrate them in typical applications.

Thus, a new field of knowledge as well as networking and orchestrating existing knowledge is opening up.

The course addresses those students from the department of electrical engineering and from the department of mechanical engineering who express an interest in interdisciplinary engineering.

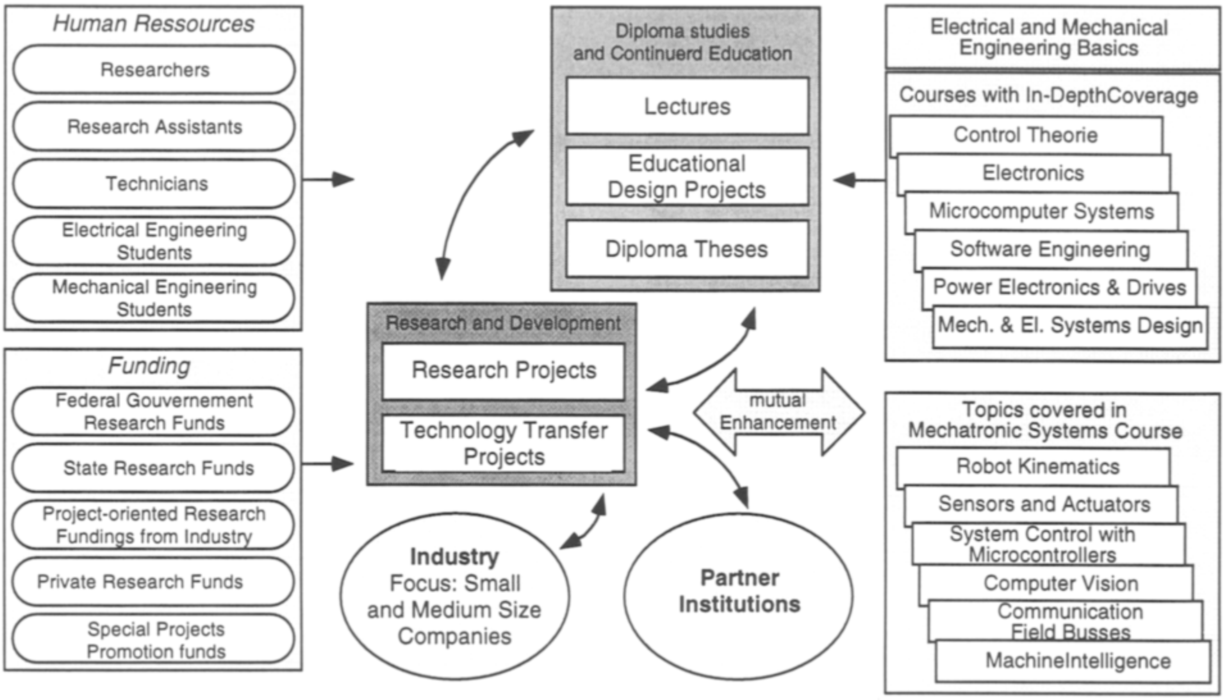

Human Ressources
Researchers
Research Assistants
Technicians
Electrical Engineering Students
Mechanical Engineering Students
Funding
Federal Gouvernement Research Funds
State Research Funds
Project-oriented Research Fundings from Industry
Private Research Funds
Special Projects Promotion funds
Diploma studies and Continuerd Education
Lectures
Educational Design Projects
Diploma Theses
Research and Development
Research Projects
Technology Transfer Projects
Industry
Focus: Small and Medium Size Companies
Partner Institutions
mutual Enhancement
Electrical and Mechanical Engineering Basics
Courses with In-DepthCoverage
Control Theorie
Electronics
Microcomputer Systems
Software Engineering
Power Electronics & Drives
Mech. & El. Systems Design
Topics covered in Mechatronic Systems Course
Robot Kinematics
Sensors and Actuators
System Control with Microcontrollers
Computer Vision
Communication Field Busses
MachineIntelligence

Figure 2. Concept for integrating education, technology transfer, research and development

The principal goals of course and the projects compounded with it s are (see also Figure 2):

- Networking and orchestrating knowledge from specialized disciplines for designing systems that otherwise could not be built.
- Teaching of essential knowledge for working in the field of mechatronics.
- Introduction to analyzing and modelling of complex mechatronic systems.
- Introduction to system integration
- Initiating contacts between students with different profiles of studies, thus creating interest and understanding for different points of view due to different curricula.
- Initiating interest and understanding for different approaches both to areas of interest and to problems in projects.
- Supporting team-work in both study and projects, and giving incentives for knowledge sharing between students.
- Providing the conditions for hands-on experience with mechtronic systems.
- Setting the basis for planning an efficient and well founded concept with a view to the further study of mechatronic subjects.
- Introducing students to projects and encouraging interest in developing projects.

5 COURSE TOPICS COVERED

The following topics are presented in theory and reinforced by means of examples and practical laboratory instruction:

- Analysis and modelling of mechatronic systems
- Robot kinematics and kinetics
- Sensors and actuators
- Application oriented introduction to microprocessor technology to control mechatronic systems.

- Communication between systems and subsystems, i.e. fieldbuses
- Principles of computer vision
- Introduction to machine intelligence
- System design and system integration

6 COURSE MATERIALS

At the beginning of each course students obtain a set of materials, which consists of:

- *Study Notes* The one semester course is based on a study note package especially developed for the purpose. The notes contain a description of the theory, application examples, practice assignments and introductions to laboratory practice.
- A *set of experimentation material* in a box. This includes: microcontroller board, breadboard, DC-motor, stepper motor, NiTi-shape memory actuator wire, multimeter, various sensors.
- An *autonomous vehicle «Pacificar»* (see Figure 3) which is used for experimenting with sensors and actuators, as well as for introduction to the programming of real-time systems.

7 COURSE DELIVERY

The course contains four lectures per week, laboratory and practice time inclusive. In addition, students are expected to spend the same amount of time working in their own time, studying and working on practice assignments. We refer to this as the „Fifty-fifty-Model“. For this purpose both class rooms and laboratories are available to students during the opening hours of the institute, which, in practice means virtually until midnight. Experimental material may also be taken home by the students

Generally we are seeking to illustrate theory continuously using adequate real-world objects. We use the following equipment which has been specifically designed for the mechatronics course:

- *SCARA robot equipped with a Computer Vision System* illustrating robot kinematics, computer vision systems, machine intelligence.
- *A sorting machine with portal robot* for the sorting of labelled blocks, and transporting the assorted objects back to a container for the continual distribution process. This is a Fischer Technique -Model, controlled via an Interbus-S field bus by a program written in C-language, running on a PC.
- A fast electro-mechanical positioning system (EMPS) with control through a signal processor, using a rapid prototyping system. This equipment serves for the introduction to servo control systems.

This equipment serves for the introduction to servo control systems.

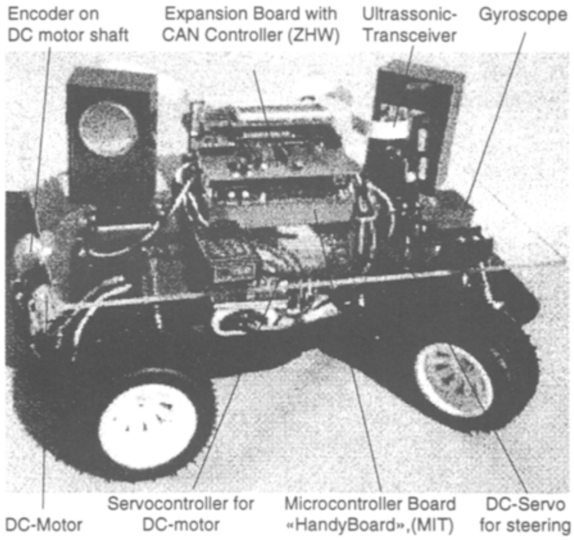

Encoder on DC motor shaft
Expansion Board with CAN Controller (ZHW)
Ultrassonic-Transceiver
Gyroscope
DC-Motor
Servocontroller for DC-motor
Microcontroller Board «HandyBoard»,(MIT)
DC-Servo for steering

Figure 3. The Pacificar experimental vehicle

8 PROJECTS

As illustrated in Figure 2 different types of projects are closely correlated. Together they represent a vital component in linking educational and research activities. This linking results in a mutual enhancement in achieving their relatives goals

- educational design projects
- diploma theses
- research projects
- and technology transfer projects

In the following examples of the different project types are given.

8.1 Design Project »Modelling of a SCARA robot in direct and inverse kinematics»

Design projects generally take place parallel with the delivery of course theory, both individually or in small groups. In the most favourable settings, the projects motivate for continuous and contextual application of acquired knowledge through appropriate selection of the problem fields.

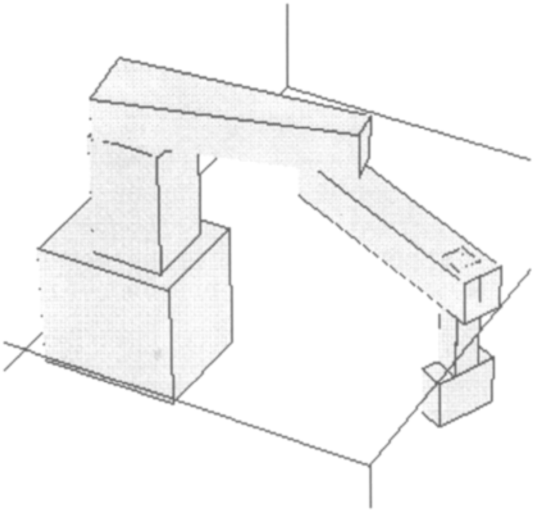


Figure 4. Example of a design project: Mathematical model of a SCARA robot

8.2 R&D Project «Expansion board»

Developers of complex robot projects often require more I/Os than the standard Handy Board[1] offers. In particular, there is often a need to control more than two servos, as is possible with the standard board. To switch from the Handy Board to another more powerful controller board is, for many developers, not desirable. This is because they do not want to despense with the many advantages that Handy Board users enjoy. Such advantages consist of the programming environment «Interactive C»[2] and the opportunity to share experience and find support through the web with a large group of users.

To meet the above mentioned requirements, an expansion board was developed for the Handy Board at the mechatronics laboratory of ZHW. The expansion board is simply plugged into the bus connectors and supply connectors on the Handy Board. Before connecting it the processor unit must have been pulled out. It is subsequently clipped into the processor extension socket on the expansion board (see Figure 5).

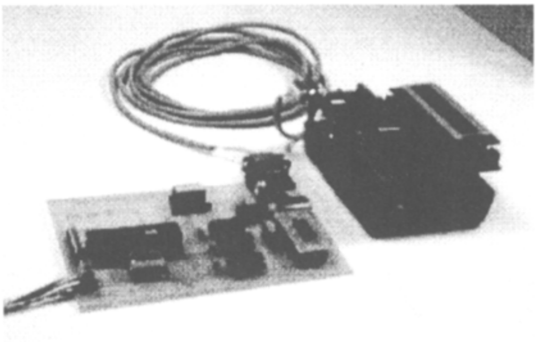


Figure 5. ZHW-Expansionboard mounted on top of the Handy Board (right hand side), prototype of a sensor/actuator module connected via a CAN bus with the expansion board.

The Expansion Board introduces the following additional features:

- 8 servo connectors for driving pulse width controlled servos.
- 10 inputs which are routed to the timer system of the 68HC11 for pulse width and frequency measurement.
- 6 non-latched I/O pins and16 configurable I/O pins.
- A CAN controller for communicating with industrial devices and with other handyboards that are equipped with an expansion board (Basic CAN 2.0A and Full CAN 2.0B).
- A multipurpose motion IC for controlling permanent magnet DC motors or brushless DC motors or stepper motors.

An additional power header with a switched mode voltage regulator for a current of 3 Ampere meets the increased energy requirement which is mainly imposed by the added servos.

There is no change in the use of the Handy Board whether it has an expansion board or not. Everything that previously worked will go on working, but with additional functions at hand.

---

[1] The Handy Board is a 68HC11-controller board based on the Motorola Microcontroller M68HC11, designed at the Massachusetts Institute of Technology (MIT) for experimental mobile robotics work. MIT has licensed the Handy Board design at no charge for educational, research, and industrial use.

[2] Interactive C is a compilation environment for Motorola 6811-based embedded systems, designed by R. Sargent and A. Wright. (Newton Research Labs and MIT)

8.3 Development Project «Modular Robot Systems»

In order to gain more flexibility in the design of multi-axes robot systems with a multiple of communicating microprocessors, we are developing a group of sensor actuator modules, each with a CAN-Bus interface. This allows us to build systems with a modular structure. In addition, this concept allows the direct integration of industrial system units that are equipped with a CAN interface. Last but not least, this would facilitate the replacement of the microcontroller Motorola 68HC11 in an existing mechatronic system with a more powerful microcontroller.

8.4 Research Project «Automatic Car Parking»

The primary research goal focuses on the transfer of theoretical fuzzy control to a practical software that is capable of controlling the process of parking a car in side entry parking space. If the process does not meet the given position standards in the first attempt, the software also includes algorithms for an incremental improvement of the final vehicle parking position. The solution of this type of parking problem is expected to be of interest to the automobil industry. The implementations of fuzzy controlled parking could be employed to assist drivers in parking cars by giving them supporting directions. Alternatively, fully automatic parking is possible. A more detailed description of the project is given in (2).

In an initial part of the project the autonomous vehicle was developed and built in a series of 14 identical models. In their most basic form the vehicles are used in practical exercises; for use as a research and development platform they are furnished with additional project-specific modules. Using ultrasound sensors the vehicles pick up signals from the surrounding area. The signals are then processed via a microcontroller which forms outgoing signals which in turn govern their actuators. In conjunction with specially developed programmes in „C“ , the vehicles are able to move about autonomously and collision free within their operational area. They are also able to fulfil tasks such as parking automatically in a given parking space, a manouvre which will be illustrated in more detail in the following text. These developments took place within the framework of special development projects and a series of interdisciplinary projects and diploma theses undertaken by students and graduates alike.

When parking a car we control the degree of movement by means of rules, estimated values and corrective algorithms, which we apply until we are satisfied with the result (or until we abandon the attempt in favour of another, and where possible larger parking space). Exiting a parking space is usually much easier than entering it. The former presents a problem of Direct Kinematics: Starting from the prescribed parking position, the vehicle is manouvred out of the space by adjusting the control entities steering angle and angle of rotation of the wheels; upon exit, the vehicle automatically aligns itself to the desired course and only has to be supervised in order to avoid collision. Conversely, entering a parking space represents a problem of Inverse Kinematics: The designated parking location is given. What is demanded is the control vector as a function of the position with which to manouvre the car into the designated space without recourse to corrective action. Furthermore, a suitable starting position has to be both determined and reached, in advance. Problems concerning Inverse Kinematics generally have more than one possible solution and determining the *control* functions presents a difficult task.

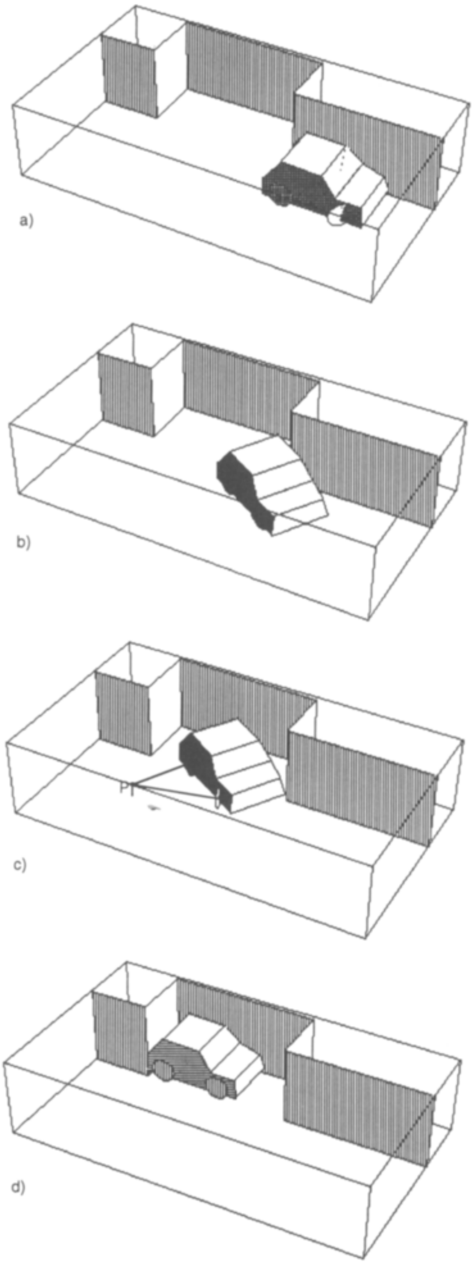


Figure 6. Four vehicle positions during the parking process along an ideal course curvature in a parking space of the smallest dimensions. From top to bottom: a) starting position, b) completion of the initial clockwise rotation, c) completion of the translation; in addition the instant center (P) for the subsequent rotation is indicated in this figure, d) vehicle in the final parking position after the second anti-clockwise rotation before centering.

In kinematic terms the parking manouvre can be described as a sequence of translatory motions and rotations. Figure 6c shows as an example one part of the rotation sequence around the instant centre (P), which is given by the intersection point of the extended wheel axis, which determines the parking process.

A parking space which is larger than the theoretic minimal gives rise to a multitude of possible courses which fulfil all the requirements necessary to allow parking without corrective adjustments. At the same time, the singular starting position is transformed into a field of possible positions from which entry-parking is effected without the need for correction. An improved course steering is thus made possible for each respective starting position. This nevertheless results in a greater complexity of steering functions.

The mathematical model illustrated in Figure 6 shows individual positions of the vehicle whilst manoeuvring into a parking space of theoretically minimal dimensions along an ideal course.

Because of non-ideal aspects of the actual process as it takes place in reality, errors in the end position can occur, In these cases, corrective algorithms are continually applied until the final parking position is achieved within the correct tolerance.

Automating the Parking Process.

In the same way that the sets of joint variables governing the steering of industrial robots is calculated, so the control vector for the entry parking process can be calculated based on the geometry of the parking space and the surroundings by means of inverse Kinematics. This deterministic solution would, however, involve difficult trajectory calculations, and would therefore for real-time control require the use of a very powerful microcontroller. In addition, measuring sensors would need to be very precise. Fuzzy Logic methods, as operational in our parking robot, are able to solve the problem more efficiently and more flexibly, including the corrective algorithms employed to improve the end position of the vehicles, even though the process may in part be inaccurately determined. A small amount of data, few rules and a simple model of reality already suffice to govern the process. An 8-Bit micro-controller with a 32 Kilobyte Random Access Memory (RAM), as is used in our experimental vehicle, is adequate to the task.

Control Concept

The demands placed upon the steering functions of autonomous vehicles are in part different from those relative to classical machine steering, a complete model of whose working parameters can be formulated. Therefore, as a rule, the processes to be carried out are completely planable, and, due to sensor data input, the outgoing steering data can be generated via an appropriate programme.

In contrast, autonomous vehicles move about in a changing and sometimes highly complex environment, and therefore cannot be entirely integrated into a model. With the Subsumption Approach, R. Brooks has introduced a concept which has proved highly valuable in many areas of use (see reference 1.).

Subsumption Approach

Subsumption Logicpresents a system of hierarchically constructed functions which work parallel to one another. These functions are referred to as „behaviors“. The lowest behaviors form the basis functions of the system; higher level behaviors refine the way the system behaves and control the special tasks the robot has to perform, as, for example, the parking manouvres in question in this paper. This structure results in robust systems. Should a higher level behavior cease to function, the lower level behaviors ensure that the system's basic functions are still available. If several of the behaviors working in parallel simultaneously give orders to the activators, one of them has to be able to suppress the others. This can be achieved by use of a so-called Suppressor Knot at the output of a behavior, or by Inhibitor Knots in the module input. In this way, for example, a link-up process has to be able to suppress a behavior responsible for collision avoidance.

Each behavior processes cyclical input data independently and produces output data. An arbitrator deals not only with areas of conflict but also with areas of priority, and establishes which of the generated data should be used and when. Output data that is not used is lost and replaced by new data.

The process pertaining to parking in a designated space is now expressed as a Finite State Machine and a model group which has been developed according to the Subsumption Approach principle.

The work done so far in this field illustrates that vehicles can be parked both automatically and autonomously. The robot thus employs the same strategies as those used when steering a vehicle manually, achieving results comparable to those of a person adept at parking.

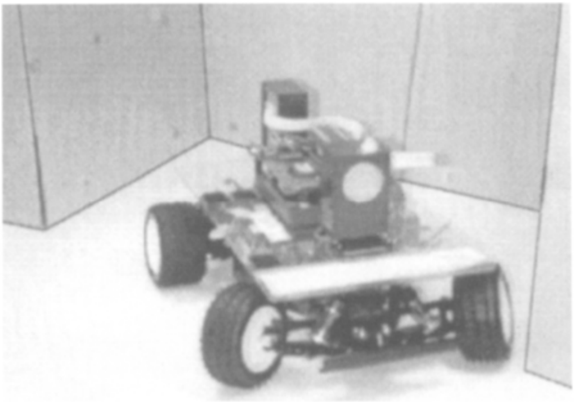


Figure 7: The experimental vehicle „Pacificar“ parking in a sideways entry parkingspace.

The use of Fuzzy Logic facilitates problem solving in conjunction with the application of a reasonable investment of hardware and software. The parking process can be made even more performance-efficient by further developing and refining the application of Fuzzy Logic steering algorithms. The improvement is demonstrated by the fact that the requirements relative to the starting position of the vehicle are less rigid, without increasing the need for corrective action.

In a practical situation, taken into consideration with a development partner from the car industry, this would mean that the task of the car driver would be reduced to that of bringing the vehicle into the parameters of a large starting field. The vehicle would then be parked by the robot in one manouvre. Alternatively, the system could be used as a parking assistant, giving the driver instructions for the parking manouvre.

8.5 Technology Transfer Project

An example of a technology transfer project is the development of the power and control electronics for a Switched Reluctance Drive (SRD). This project is carried out in a close partnership between a small company for designing and manufacturing electronic drive and control units for consumer and industrial applications and the mechatronics laboratory, where the mechatronics laboratory provides the special know-how in power electronics and control with microprocesseors, as well as the infrastructure for developing and testing.

9 CONCLUSIONS

In this paper, we have presented a concept for linking educational and research activities The fact that different types of projects are closely correlated with lecturing constitutes an important element within this concept. In particular, this is illustrated by the Pacificar project. As a whole the different project types form a vital component which contributes substantially to meeting the three main obligations, i.e. lecturing, technology transfer and research.

With the introduction of the described concept an significant build up of research and development competence has been achieved. Students get in an early stage involved with research and technology projects and many get motivated to engage themselves more deeply in such projects. For young researchers and developers this provides an interesting opportunity to widen and deepen their knowledge in mechatronics.

10 ACKNOWLEDGEMENTS

The author wishes to acknowledge the contributions made by undergraduate students through their design projects and diploma theses relative to the Pacificar Procjet. I wish to extend thanks to both the research assistants and the staff of the mechatronics laboratory for their indispensable contributions to the project.

REFERENCES

1. Rodney A. Brooks, A robust layered control system for a mobile robot, IEEE Journal of Robotics and Automation, Vol. RA-2, No. 1, March 1986

2. Charles Brom, Projekt PACIFICAR: Automat zum Parken von Automobilen, TECHInfo 2/1998, Zürcher Hochschule Winterthur

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

MECHATRONICS EDUCATION AT ETH ZURICH BASED ON 'HANDS ON EXPERIENCE'

Roland Y. Siegwart[1], Roland Büchi[2], Philipp Bühler[2]

[1]Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland
roland.siegwart@epfl.ch
[2]Swiss Federal Institute of Technology Zurich (ETH), Zurich, Switzerland
buechi@ifr.mavt.ethz.ch, buehler@ifr.mavt.ethz.ch

ABSTRACT:

The ETH Zurich has a long tradition in mechatronics education. One successful part in mechatronics education is a lecture called *smart product design* with a main emphasis on *hands on experience.* Apart of the lecture giving the basics for mechatronics, the students are working on a real mobile robot system during the course. Every student team gets a mobile robot kit, based on a Motorola 68332 micro controller. It runs the real time operating system XOBERON, based on Oberon (object oriented successor of Modula2 and Pascal).

Each year's course ends with a competition between the *fully autonomous* robots of the interdisciplinary student teams. In last year's competition the students had to build an autonomous vacuum cleaning robot.

In our paper the lecture and hardware will be described and experience gained in our course during the last six years will be discussed.

1 INTRODUCTION

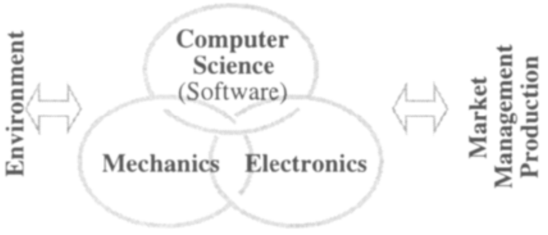


Figure 1. Mechatronics integrates mechanics, electronics and computer science to an 'intelligent' system

Most of today's engineering products make use of some kind of '*intelligence*'. This intelligence makes a simple product a mechatronics product (fig. 1). Therefore, today's engineers need a basic knowledge on mechatronics design techniques and design elements such as sensors, actuators, mechanics, electronics, and software [2]. However, to acquire this system knowledge, building complete mechatronics products are a necessity. Creativity, competition, marketing and teamwork are very important issues for innovation and has therefore also to be addressed in engineering education.

In 1992 the course *smart product design* was initiated at the Swiss Federal Institute of Technology (ETH) to support the students' understanding in mechatronics systems design. A number of smartROB design kits have been built and have been in use since. They are the basic platform for a student contest in mechatronics design, within the framework of the graduate and post-graduate course *smart product design.*

2. EDUCATIONAL CONCEPT

Mechatronics education at ETH Zurich covers robotics, kinematics and dynamics, control, microprocessor systems, real time programming, production techniques, computer vision, sensors and actuators and much more [1]. In a successful mechatronics design all these key technologies mentioned above have to be integrated to an intelligent system. Through a whole variety of lectures the students from mechanical and electrical engineering and computer science get a profound knowledge in mechatronics. However, the *design and building* aspects, which are very crucial for mechatronics education, are not covered in the basic lectures. Therefore the lecture *smart product design* was created with its main focus on *hands on experience* and *learning by doing.* In the theoretical part, the basic for successful and innovative mechatronics design is taught. The laboratory work based on the smartROB design kit, enables the students to use their theoretical knowledge and to build a winning mechatronics system.

The lecture and the laboratory work are run parallel, usually in the summer semester, from April to July. Even if the course gives only 3 credit, the students usually work around 100 to 200 hours on the project. They usually attend the course one or two years before their final diploma (equivalent to a master's degree).

During the first week of the course, the students are invited to form their teams by themselves. Usually teams consist of students from different departments. However, teams with all students from the same department are also accepted. The available laboratory space can cover 8 teams with 4 to 6 students each.

2.1 Primary goals of the course

To be successful in mechatronics projects, students have to learn to think interdisciplinary and to work in an interdisciplinary team. Within the course, they are forced to cope with new disciplines, to communicate with other engineers and to integrate all to an optimal system. They have to learn about the terminology of other disciplines, to specify the system and subsystems and to define hardware and software interfaces. Within the project they also get an idea what system integration means, only by doing. Students learn how much time it costs to bring components, built by different people from different disciplines, together and experience that there are always some details not considered in the first step. They acquire the experience what it means to finish a project in time with all the technical surprises ("Murphy's law") and human factors.

Mechatronics design means also overcoming the fear of other disciplines. Therefore electrical engineers are encouraged to build mechanics, mechanical engineers to write software, and students from computer science to build electronics, and vice versa.

In the laboratory work students get a better knowledge of their own abilities and they have the possibility to develop creative ideas and to realize them.

Besides the technical aspects, we want the students to think and to learn about their social and communication abilities within a team.

All these aspects mentioned above are probably more important for successful work in industry than the pure theoretical background. They can not be covered by conventional courses, but can easily be addressed on real project work done in the lab.

The course is based on a mechatronics project because nearly all of today and tomorrow's products are based on different engineering disciplines. In the course students have to build fully autonomous systems. The realized robot has to solve unforeseen problems by itself, which can only be realized by introducing some sort of primitive 'intelligence' to the system. This makes the students sensitive to important aspects like reliability and robustness, and reminds them how small the technical abilities are compared to the abilities of humans.

Finally, building an autonomous robot and presenting it in a contest is very attractive to the students and motivates them to enormous efforts.

2.2 Lecture

The main goal of the lecture is to give the basic knowledge for the laboratory work, but also to motivate the students for innovation. Within the course we do not teach much new things, but we try to remind the students of what they already know from all the other courses. Important parts of the lecture are:

- Basics of the smartROB kit
  - electronics, mechanics, modeling
  - available sensors and actuators
  - real time environment, real time programming

Since the laboratory work starts at the beginning of the course, we have to teach this part first.

- Elements of Mechatronics Systems
  - physical effects and its application for sensors and actuators
  - basics on electrical circuit design
  - sensors and actuators

This gives a short summary of the theoretical knowledge that could be useful for building own sensor and actuator systems within the course. We expect them to remember what they learned in other courses and to realize that they can start to apply this knowledge now. In the lecture we lay stress on understanding the physics of sensors and actuators, before applying it to the system.

- Control
  - architecture and design of control
  - linear control, behavior based control, fuzzy control, neural networks
  - microprocessor systems, digital controller hardware

As control theory is an abstract and highly mathematical field, many students have difficulties to get the link to the 'real world'. We show how to design a model based control and how to implement it on the smartROB, based on the example of velocity control. During the course they have to implement all kinds of controllers, for example low level speed control, path planing or global strategy.

- Mechatronics Design
  - integrated product design
  - systems engineering

An example, how to structure a mechatronics project and how to specify the different task and interfaces is presented. We want the students to realize conscious that they are doing such kind of design right now.

- Creativity / Innovation
  - What is creativity and how can we improve it
  - working in teams, evaluation of teams
  - from the idea to the product

We remind the young engineers that there are persons behind each innovation. Usually various non technical problems have to be solved besides the technical ones, for example personal problems: how to motivate themselves and the team if the project gets stuck. Students learn about the importance of close cooperation within the team. They get the chance to evaluate their teams using a standard test for the personal style of learning and working.

2.3 Laboratory work

Two laboratory rooms are available for the course. Each student team gets a smartROB kit, a Macintosh computer, some additional sensors and actuators, a laboratory table and a cupboard with tools. Additional electrical and mechanical hardware is obtainable on request. The lab is also equipped with a small lathe, a small milling machine, a drill and a bench for mechanical work. In one of the labs the play-field for the championship is at least partly installed. The students have free access (day and night) to the lab throughout the whole semester.

In the first 2 to 3 weeks they have to learn about the smartROB kit. In the second week they have to fulfill an open loop task with the smartROB. In the third week they have to program their smartROB to follow a wall avoiding various obstacles using sensor feedback control. Already for there first two tasks the robots have to operate *fully autonomously*.

After the third week, the teams are absolutely free to continue their work according to their own project plan. The only fixed time is the championship, that is about 10 weeks later.

Half a day a week is reserved to discuss problems with teaching assistants and staff from the Institute of Robotics. Twice during the course each group has an official meeting with the lecturer. This allows us to follow up with the actual state of the project of each team and to give some important tips and hints.

2.4 Experience in the last six years

From the very beginning, the course *Smart Product Design* was very successful and appreciated by the students. It demonstrated the need of such alternative educational concepts and motivated us to continue.

Since many students have no practical experience and since they sometimes work deep into night, all that is handed over to the students has to be very robust (or inexpensive). Everything that could probably break will for sure break if a student team is working on it. However, our current smartROB kit became '*student proven*' during the last years and we are very positive that in the future it can withstand all attacks from enthusiastic and inexperienced young engineers.

We and also the students realized that the timing and early testing are very essential factors for a successful design and good performance at the championship. All the finalist of the last years had an excellent timing during the course. They started implementation and testing in a very early stage and have therefore been able to identify and solve all problems before the championship.

Robust and relatively stupid machines often turn out to behave better than too sophisticated solutions, because testing and optimization of sophisticated machines are very difficult and there will always remain some unforeseen problems.

The students also get some experience about the feasibility of their own ideas and how to organize and to work within a team. For us it is an excellent opportunity to learn about the capabilities and potential of the individual students. Most of our doctoral students started as successful young engineers during the smartROB course.

3. THE CONTESTS

The smartROB course always ends with a contest, the smartROB championship. That might rise the question, if there is really a need for a contest? Our experience shows us that students like to present their product to a large audience and to demonstrate its performance in competition to others.

The smartROB championships are carried out in the big hall of the main building at ETH in the last week of the semester. The hall allows spectators to look down from three floors on the play field and gives an excellent ambiance. The championship is usually broadcasted by the local radio and television stations and has a wide response in the newspapers.

However, the contest itself is not taken too seriously by the students and they have a lot of fun.

The final mark we have to give to the students for their diploma (master's degree) is not dependent on the result of this contest only, but is also based on the creativity, the hardware concept, and the organization within their group. In addition we test each student separately during a 30 minutes oral exam.

3.1 Review of the last years contests

The task of the first year's championship in 1992 was to find golf-balls on a four by four meter field and to place them in a center hole.

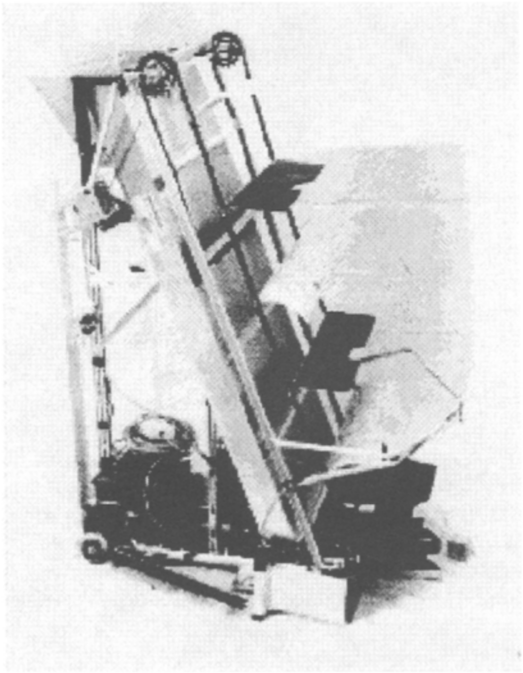


Figure 2. Garfield the second finalists of the championship 93.

For the 1993 championship the students had to design and fabricate a fully autonomous system building a tower out of wooden blocks (fig. 2). The highest towers built by the autonomous robot systems were more than 1 meter high. The winning team focused on very high speed. They developed a special interrupt procedure to accelerate stepping motors to much higher speed than the standard TPU software had allowed.

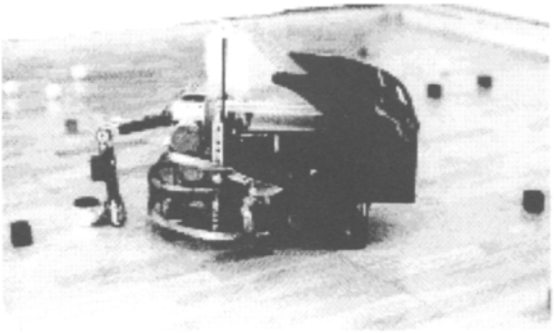


Figure 3. Pygmalion, the winner of 94's final. It perfectly collected and identified the different blocks, no industrial product could have done better.

In the 1994 championship, the robots had to collect different objects in a 8 by 8 meter labyrinth (fig. 3). The object's weight or color had to be detected by the robot system. The winning team presented a nearly perfect robot system, with robust sensors for navigation and identification of the objects, an excellent mechanical hardware design and impressing maneuvering abilities.

In the winter semester 1994/95 there was a special course with only 3 teams. Each team got two smartROB kits to solve a transport problem. The winning team designed and trained their robots for optimal cooperation. For the first time also one team of an other engineering school was present.

In the summer 1995 event, the robots had to play soccer. Twenty green and red balls were distributed on the 8 by 8 meter play field. The task of the two robots on the field was to kick the red balls into the 40 cm wide goal of the other team and vice versa. The design varied from robots able to collect and kick single balls to robots with a big propeller blowing all balls towards the goal of the opponent.

In summer 1996, the robots had to find burning candles in a 8 by 8 meter labyrinth (similar to the 1994 championship) and cover them with a small metal cup until they were extinguished.

The competition during the winter semester 1996/97 was focused on learning. The robots had to explore a labyrinth and to learn the fastest way through it.

The goal of most recent event in summer 1997 was to develop an autonomous vacuum cleaning robot. During the competition the robots had to clean a surface of 8 by 8 meter covered with little wood pieces, tables and chairs. The solutions of the 8 student teams covered the entire spectrum of possible approaches, some more oriented toward sophisticated mechanical design and some more towards precise navigation and obstacle avoidance.

4. BASIC HARDWARE OF THE SMARTROB

To allow the students to build a sophisticated robot system within a limited amount of time, an appropriate basic hardware is of high importance. Therefore our basic smartROB design kit is described in the following.

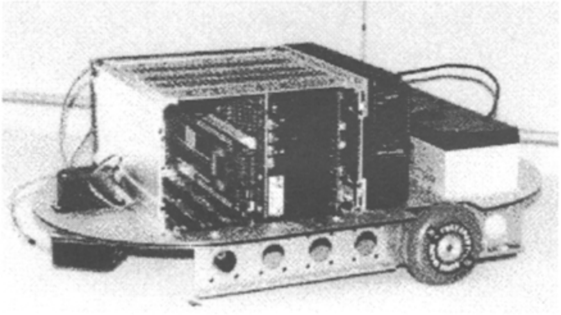


Figure 4. The new basic smartROB platform with electronics rack, two batteries, range sensor and differential drive with two DC motors. Its length is around 45 cm.

4.1 Electronics

The goal of the basic hardware is to provide a cheap and flexible mechatronics design kit, containing design elements such as a micro controller, different sensors and actuators, and mechanical hardware. As a result, the smartROB design kit has been developed (fig. 4 and 5), with a Motorola 68332 single chip processor as main controller. It plugs on an interface card, providing various analog and digital input/output channels and four incremental encoders. A power card allows to control three stepping motors, two DC-motors and four servo-motors. One ultrasonic and one infrared distance measurement system is included in the kit. Other sensors and actuators can be added by the student teams using free analog and digital i/o's. The whole electronics is assembled in a small rack (fig. 5) including the power supply for 5V and ±15V. Two rechargeable 12V batteries, connected in series, are used as energy source. They allow autonomous operation of the smartROB for around half an hour.

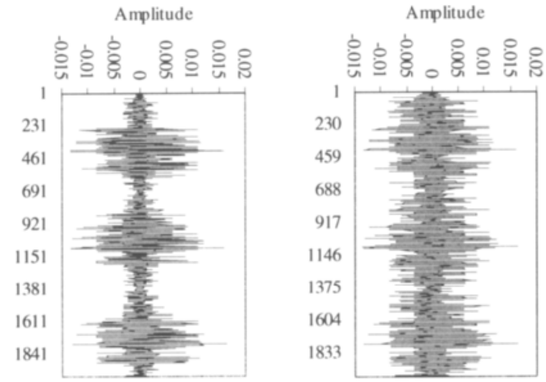


Figure 5. Schematic of the smartROB electronics consisting of the 68332 business card computer and various boards for sensor and actuator control. The optionCard is a free slot for sensor and actuator electronics developed by the student teams.

As mentioned, the 68332 processor has a time processing unit (TPU) allowing for stepping motor control and pulse width modulation. The powerCard makes extended use of this nice feature. Twelve TPU channels are used to drive three stepping motors (four channels for each motor). The remaining four signals of the TPU are used to control servo motors, the type of motors used for remote controlled toys.

To get the student teams to build their own sensor electronics, the task of the contest always requires some additional sensors. The additional electronics is built up on so called extensionCards. At the rear end of these cards all analog and digital inputs and outputs including are available on a predefined 96-pin connector. The rest of the card is free. Students can use wire-wraps or other types of breadboards or even their own printed circuit boards.

4.2 Mechanics

The basic mechanical hardware of the smartROB (fig. 4) is a mobile platform with three wheels. The chassis is based on aluminum profiles. The two rear wheels are driven by two independently controlled DC-motors with integrated gear box. For optimal position and speed control, the motors are equipped with encoders. However, usually the mechanical hardware is totally changed by the student teams during the course. For mechanical changes and extension of the platform additional aluminum profiles, aluminum parts from hobby shops, gears, pulleys, wheels and much more is available.

5. SOFTWARE ENVIRONMENT

The 68332 micro controller runs the real-time operating system XOBERON developed at the Institute of Robotics and is programmed in the high-level language Oberon. Oberon is an object oriented successor of Modula2 and Pascal [3]. Alongside, a software library for handling a wide range of processor functions, I/O, basic sensor programming, and robot motion functions is provided. Due to the modularity of the system, the compatibility of its components and the ease of program development and testing, various configurations of mechatronic products can thus be built and tested quickly.

XOBERON is a cross development system running on Macintosh, Sun and PC as host computer. It supports processors of the Motorola 68-thousand family and PowerPC. The target system is connected through a serial link with the host computer (fig. 6). A Monitor program that executes basic commands, e.g. for downloading the program and debugging, runs on the target processor.

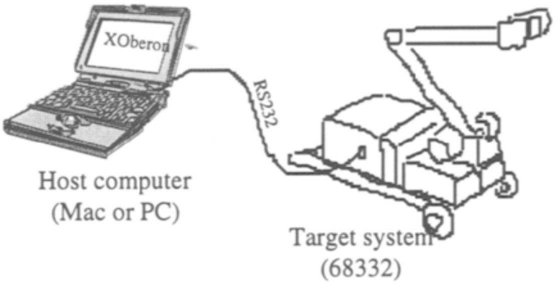


Figure 6. Real time software environment. The software developed on the host computer can easily be downloaded and debugged through the serial line.

The most important features are:

- the target processor uses a window on the host as a terminal
- incremental linking allows to add and remove tasks without interrupting the other running tasks.

The XOBERON system can easily be operated without a prior knowledge in real time operation systems. Students can start to work within about one hour.

6. SUMMARY AND OUTLOOK

A different approach of mechatronics education, based on 'learning by doing' has been described. Based on the smartROB, a basic robot design kit, students apply their theoretical knowledge on real hardware. By adding their own sensors and actuators and by writing 'intelligent' real time software, the students are able to design and build an autonomous mechatronics system. Each year's lecture ends with a competition between the interdisciplinary student teams. Our experience of the last 5 years shows the importance of this type of lectures. It helps the students to get a deep understanding of the theory and motivates them to apply their knowledge.

Feedback from the students of the last years and our personal experience encourages us to continue with the course and to improve it each year, even if the preparation is a pretty hard job.

7. ACKNOWLEDGEMENTS

The described education project was possible thanks to enormous support of the whole Institute of Robotics. Special thanks go to Professor Gerhard Schweitzer, leader of the Institute of Robotics and to our people in the electronics lab and the workshop. High appreciation goes also to all the students of past years who, through their enthusiasm and creativity, made this new education concept a very successful event.

REFERENCES

[1] G. Schweitzer, Mechatronics - Basics, Objectives, Examples, Proc. Instn. Mech. Engrs., IMechE, Vol. 210: 1-11, 1996

[2] E. Carryer (editor), Workshop on Mechatronics Education, Stanford University, July 21-22, 1994

[3] D. Diez D. and S. Vestli, D'NIA, an Object Oriented Real-Time System, Real-Time Magazine, Real-Time Consult Publication, Brussels, August 1996

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronics Education in Unified Curriculum

- case study of CNU -

Cheol-Woo Jo[a], Soo-Tae Kim[b], Chee-Ryong Joe[c]

NSSPC,
#9 Sarimdong, Changwon National University, Changwon, Kyeongnam, 641-773, Korea
[a] cwjo@sarim.changwon.ac.kr, [b] stkim@sarim.changown.ac.kr, [c] jcr@sarim.changwon.ac.kr

This paper describes the mechatronics education system in the College of Engineering, Changwon National University, Changwon, Korea. The system is peculiar in that five different departments are combined together and established mechatronics core course and five different major courses. New laboratories are built to educate the core courses including mechatronics related experiments. This reformation of the engineering education system became possible based on a special fund given by the Korean government.

1. Introduction

These days engineering technology is changing rapidly and the changes in the engineering education system is required.

Mechatronics has already been emerged as a new engineering field. But the method how to educate the students to cope with the field has not been yet established. Formerly different areas, for example mechanical engineering, electrical and electronic engineering, were taught separately. Mechanical engineers were hard to understand microprocessor systems or electronical parts, and electronic or electrical engineers were hard to understand mechanical systems. Since most of the machines in the field are controlled by microprocessors, both the knowledge of mechanical and electronic engineering are essential to run the machines or to develop new ones. Thus the engineers were trained again in the company to understand the system in the field. So there has always been needs of practical education, which can be applied directly in the field. It was one of the main reasons for the Changwon National University to reform the education system.

In this paper, several aspects of the changes in our educational system are discussed.

2. Motivation

The direct motivation of this activity was the financial supports from the government. The Ministry of Education of Korean government tried to promote the reform by selecting 8 universities throughout the country, and giving them special fund for the activity. Proposals were submitted by almost all the universities. Each universities were requested to specify a field, which would be reformed. After a serious evaluation 8 universities were chosen and each of them were to be funded 5 billion won(approx. over 5 million US dollors equivalent at that time) per year for the 5 years from 1994. The college of engineering of Changwon National University was chosen to be specialized in mechatronics.

The main reason to have chosen the field of mechatronics as our focus was that our university was located inside the Korea largest mechatronics industrial complex. Biggest mechatronic companies were located near the university and school-industry cooperation and field education was very easy. No other university in Korea had better environment than CNU.

3. Educational System

3.1 Mechatronics Core Structure

The core structure of mechatronics was built by combining three major areas such as mechanical engineering, electrical engineering, electronic engineering. By combining these three different areas we can expect synergical effects in understanding and implementing complex mechanisms, which consists of most of the major modern products. Even though mechanical system forms the main structures of many products, the electronic and electrical systems, such as microprocessors, servo motors, controls etc moves the parts in the product.

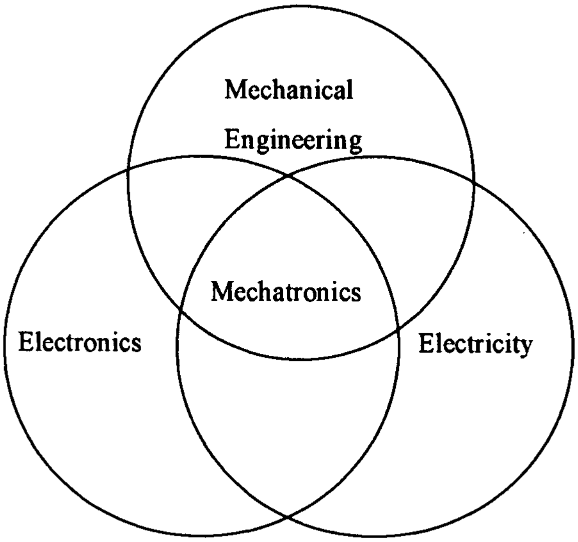


Figure 1 Concept of Mechatronics

Figure 1. is a well-known diagram of the structure of mechatronics.

3.2 Interdisciplinary Curriculum

In order to achieve the goal of mechatronics education, we combined five heterogenious departments into one group of departments, i.e. mechatronics engineering division. The first and the most important barrier against the combination was the agreements of the faculty members in each departments. Every faculties agreed to the idea of mechatronics but disagreed to the mixture of the departments. In Korean system such inter-discilplinary education is hard to achieve in general.

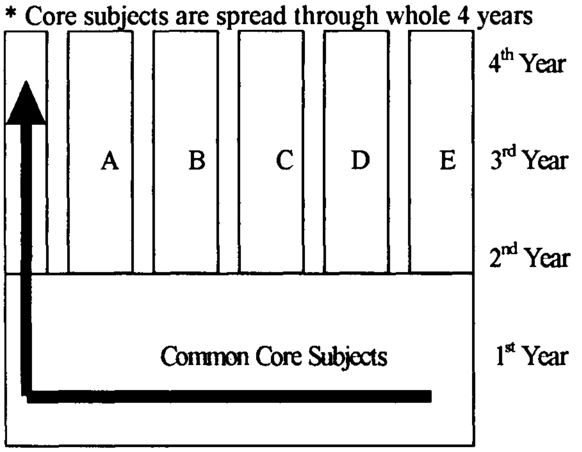


Figure 2 Structure of Unified Curriculum.

Three possibilities were considered. Those are; 1. Combination of the departments, 2. Combination of the curriculum, but separate departments, 3. Independent depatment and curriculum. After hours of serious discussions between the faculties, we reached a conclusion, unification in curriculum level maintaining individual departments. First, core subjects of the mechatronics were extracted from existing courses or new subjects were created. Table1. Shows the core subjects of the mechatronics education. These subjects includes mechanical, electronic, electrical subjects, which can be considered to be a core of mechatronics.

Students learn total 150 credits during their four years including some practical subjects such as etiquettes in industrial works, practice in industry. Figure 2 shows the structure of the unified curriculum.

In their first year, all the students of the machatronics group learn the same basic core subjects. These include basic electronics and electricity in theory and experiments, basic mechanics in theory and experiments etc.From their second year, they learn subject for each departments as well as core subjects. Each subjects can be taught in different semesters depending on departments..

All the students of the mechatronics engineering division are requested to take all the common subjects of table 1. Some of the common subjects are offered from their 2nd year. As an option they can choose 9 subjects among 12. In addition to mechatronics core subjects they learn specific subjects for their own departments. Student choose their departments at the end of their 1st year. From the 2nd year they learn their own specific subjects as well as remaining core subjects.

2.3 Practical Education

To make the students have practical experience in work place all the students who belongs to the machatronics group are requested to take factory practice at their 1st semester of 4th year. All the students are requested to choose one company from the list of companies who are willing to accept the students as a practicer. Various kinds of companies volunteered to accept our students. This was possible because our university is located inside the largest mechatronics factory area. The students performs the field practice at the factory for one semester.

Formerly each department had their own laboratories and their education was limited to their own area with different contents. Now the students can do the same experiments at the same laboratory while each department can give more specific experimental subjects to their students. Teaching assistants are assigned to all the core experiments.

To promote the practical experience we started a mecharonics design contest. The students who want to participate in the contest submit their proposals by december. If selected, the project is funded by the school and the students submit their results (reports and workouts) by the next november and exhibitions of the works are open. This(the first) year theme of the contest is ROBOT GOLF.

3. Conclusions

In this paper we introduced our mechatronics engineering division. The contents are very similar to the mechatronics education in other institutions in the world. Masjor difference is at the unification of the five different curriculums concentrating on the

Table 1 Core subjects of the Mechatronics engineering division

Common subjects		
Physics & Experiments	Linear Algebra	Chemistry
Computer Programming	Computer Practice	Engineering Math
Machanics Area	**Electrical Area**	**Electronics Area**
Statics	Electrical Circuit	Electronic Circuit
Dynamics	Electromagnetics	
Intro. To Mechanics		
Optional (Choose 9/12)		
Hydraulic Systems	Automatic Control	Microprocessors
Robotics	Power Electronics	Signals and Systems
Production System	Servo Systems	Measurements
Mechanical CAD		
CAM		
Energy Systems		
Experimental Subjects		
Basic Mechatronics Experiment	Automatic Control Experiment	Microprocessor Experiment
Automation Experiment	Basic Electricity & Electronics Experiment	Measurement Experiment
CAD/CAM Experiment		

mechatronics core subject. We have launched our system just 4 years ago and no graduates yet. The fruit of our extremely experimental education system will be judged by the industries after employing our students.

Before launching our system we referenced many cases from european and US universities. We knew that no system can fit to anothers perfectly. So we designed our own. We would like to thank many institutions who helped us at the first stage.

References

1. Project for the renovation of education and engineering on engineering, CNU, 1994

2. Report on the NSSP, NSSPC, CNU, 1997

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

TESLA - A modular system for laboratory automation

Michael Robrecht

Mechatronics Laboratory Paderborn, University of Paderborn,
Pohlweg 55, D-33098 Paderborn
Phone: (++49) 5251 60-2421, Fax: (++49) 5251 60-3550
E-Mail: Michael.Robrecht@mlap.uni-paderborn.de
Internet: http://MLaP.Uni-Paderborn.de/

In the efficient development of mechatronic products, a vital factor - along with factors like quality and innovation - is the time interval between the initial idea and the series manufacture of the respective product. In order to meet the demand for short development times, it is essential to detect and correct errors already at an early stage of development. This is one reason why the TESLA system (Test Site for Laboratory Automation) was elaborated at Mechatronics Laboratory Paderborn (MLaP). It is meant to support the engineer in the systematic analysis of mechatronic products.

1. INTRODUCTION

The TESLA system was developed in cooperation with Siemens Nixdorf Informationssysteme AG (SNI). The aim of the cooperation is to improve the dynamic behaviour of complex mechatronic systems in precision mechanics, such as automatic teller machines, passbook printers, and document printers.

A characteristic feature of complex mechatronic systems is the strong interconnection between mechanical components, information processing, and a certain number of actuators and sensors. In order to make these components interact smoothly, extensive tests are usually required. A system to test mechatronic systems should therefore comprise not only corresponding measurement functions, but also allow to excite the test object. TESLA meets all these requirements due to its hardware and software components for signal generation and signal recording.

TESLA enables the user to test a novel mechatronic product already at an early development stage and thus to give both the construction engineer and the firmware developer detailed indications as to the system behaviour. Along with the increasing demands on the products to be tested by TESLA, ever higher demands are being made on TESLA itself. Since its first presentation [1] the system has been extended by important functionalities. After a short survey of the former state of affairs, the paper will present these novelties.

2. FORMER STATE OF AFFAIRS

Figure 1 displays the former TESLA-System. Its most important components are the graphical user interface, a transputer software, and the hardware box.

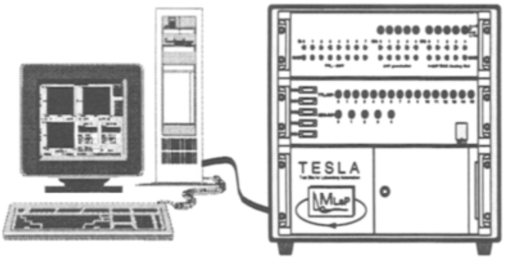


Figure 1 - Former TESLA system

2.1. TESLA Hardware Box

The control panel is the Transputer-Interface-Board (TIB) [2] developed by MLaP that can be equipped with up to four transputer modules (TRAMs) [3]. The TIB disposes of serial link interfaces that couple it with further transputer boards. Thus the computing power of TESLA becomes scaleable. To connect the TIB to peripheral boards such as ADC or DAC boards a PHS-Bus interface was integrated into the TIB [4]. TESLA employs it to control an ADC, DAC, Digital-I/O, and encoder board, all of them by dSPACE [4].

To operate DC motors there are five servo amplifiers (± 36 V/8 A). The servo amplifiers are controlled by the DAC board. Current-source electronics with 16 controlled current sources (0...64 mA) are used for opto-electronic elements. The four stepper-motor power electronics (36 V/ 4 A) plus a controller were developed for printer tests. A special feature of the stepper-motor electronics is the combination of the 7-bit micro-step operating mode that can be configu-red by software and the maximum current consump-tion that can also be determined by software.

2.2. TESLA Software Components

The software is divided into a graphical user interface (GUI), representing the interface between the user and the system, and a transputer software to control the hardware. The user interface was developed under LabWindowsCVI [5] and runs on a PC. Figure 2 displays TESLA's main GUI element. The most important moduls of the user interface are the following:

- operation of the hardware components
- visualization of measurement data
- time-domain analysis
- frequency-domain analysis

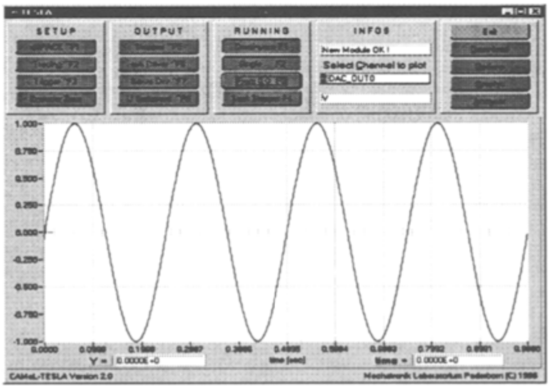


Figure 2 – Main GUI element of TESLA

The commands generated from the interface are transmitted to the transputer software. The transputer software is split up into processes working in parallel.

3. NOVEL FUNCTIONS AND MODULES

In order to make the TESLA system able to meet future demands, we had to build it up in a modular way. This regards both software and hardware. Thus it has become far easier to integrate novel functions into the existing system. Figure 3 displays the novel TESLA system:

Figure 3 - New TESLA system

3.1. Novel Software Modules

As regards novel software functions, the user has the following modules at his disposal:

- operating control
- modal-analysis function
- flexible stepper-motor control
- hardware-in-the-loop simulation (HIL)

Before expounding the functionality of the new software modules, the following chapter presents a survey of the resulting structure of the TESLA software.

3.1.1. Software Structure

Figure 4 shows the above-mentioned division of the software into a graphical user interface and a trans-puter software. The user interface comprises not only the visualization of the measured data, but also the modules for time-domain, frequency-domain, and modal analysis. Other modules coordinate the operating control, the stepper-motor control to generate the actuator sequences of the steppers, and the HIL simulation.

The transputer software has been split up into processes working in parallel. These are distributed to the three TRAMs with which the TIB has been equipped. The software was written using the Inmos ANSI-C toolset [6]. The command interpreter on the Root-TRAM evaluates the commands emitted by the user interface and initiates the respective actions of the hardware. The measurement-data management has the task of coordinating TESLA's recording measurement data. Part of it are the collecting of data generated by the processes described in the following and the transmission of these data to the user interface. The stepper control serves to transmit the stepper-motor control signals generated in the user interface to the stepper-motor electronics connected to the control. The processes on the „dSPACE TRAM" coordinate the actuating of the dSPACE peripheral boards and thus the HIL simulation.

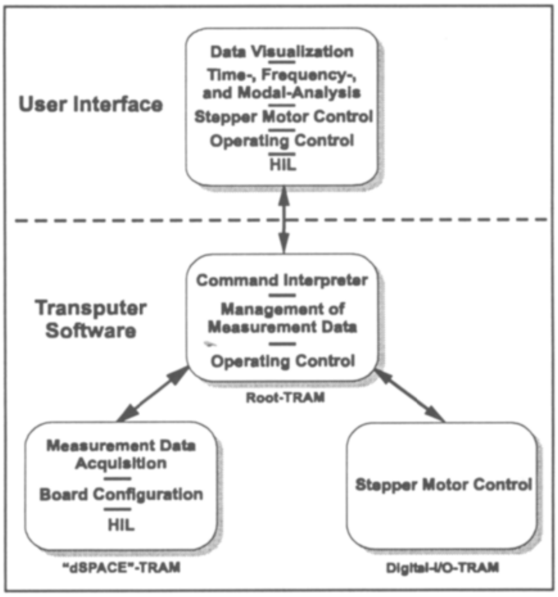

User Interface
Data Visualization
Time-, Frequency-, and Modal-Analysis
Stepper Motor Control
Operating Control
HIL
Transputer Software
Command Interpreter
Management of Measurement Data
Operating Control
Root-TRAM
Measurement Data Acquisition
Board Configuration
HIL
"dSPACE"-TRAM
Stepper Motor Control
Digital-I/O-TRAM

Figure 4 – TESLA software structure

3.1.2. Operation Control

Operation of the TESLA hardware can be effected either by mouseclick on the corresponding buttons of the graphical user interface or by means of an operating control. The latter is currently being worked upon. Figure 5 displays its concept.

A script written by the user is at first processed by a script interpreter integrated into the user interface. The script may consist of individual commands, e.g. reading of an ADC input, or comprise complex command sequences for testing teller machines. At present, the script is written with the help of a text editor. For future development, we are projecting a graphical script production.

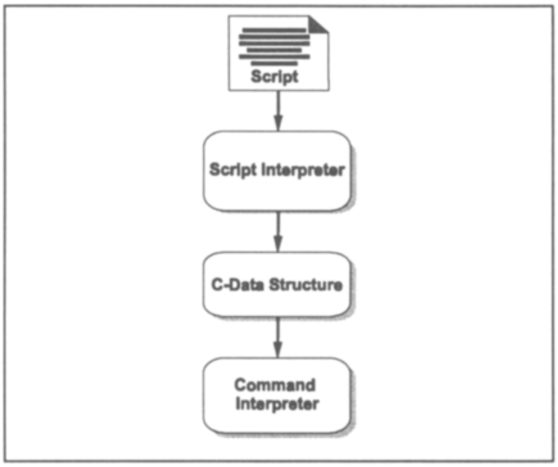

Script
Script Interpreter
C-Data Structure
Command Interpreter

Figure 5 – Concept of operating control

For realizing the script interpreter integrated into the user interface, there is the interpreter language Smalltalk. Compared to the compiler language C, an interpreter language like Smalltalk offers the advantage of allowing to employ already existing scanner and parser functionalities when processing a script. For an exchange of script data between user interface and transputer software, an interface will have to be made up because there is no approprite interpreter available for the transputer. The interface consists of C data structures. A simple "interpreter" written in C was implemented in the transputer software. It analyzes the data structures and executes the corresponding actions of the TESLA hardware.

3.1.3. Modal-Analysis

The modal-analysis function extends the former TESLA analysis features by a possible vibration analysis of mechatronic systems. The three analysis methods integrated in TESLA make use of the same measurement data and vary only in the way they process and visualize these data. The modal analysis is employed to determine the eigenfrequencies, the modal damping, and the shapes of eigenoscillation of the structure in view. These data then serve to display the eigenoscillations in their geometry and amplitude relationship.

For this purpose we implemented in TESLA diverse algorithms for the identification of modal parameters from measured frequency responses [7]. For a subsequent visualization of these data, we put to use a simple graphics editor and an animation module. Besides, we created an interface for a data exchange with the FFT analyser. The user employs the graphic editor to make a graphical representation of the structure he wants to examine. This representation will also contain the measuring points for the modal analysis. With the animation module, the user can then animate the graphical representation.

3.1.4. Stepper Control Software

The stepper-motor excitation can excite up to four stepper motors quasi in parallel. To every motor may be assigned a ramp and a trajectory. The ramp indicates the time intervals between the motor steps. The trajectory contains information on the number of steps, direction of rotation, acceleration, micro-step mode, and motor current.

3.1.5. Hardware-in-the-Loop Simulation (HIL)

The transition from a purely simulated mechatronic system to one that is realized entirely in its physical shape is often very demanding and error-prone. The HIL simulation offers the opportunity to do the transition in steps. In such a simulation, for instance the mechanical, electrical, or hydraulic components of a system are replaced by real-time simulation computers and linked to real system components in a closed control loop. TESLA's HIL functionalities are based on the opportunity to employ the model description languages in use at MLaP [8] to generate C code [9] that can then be downloaded on the TIB. Thus those partial components of a novel system that do not exist as hardware can be simulated by means of TESLA.

3.2. Hardware

Hardware build-up was completely redesigned. The hardware was divided into three modules (see Figure 3). The "basic module" comprises the TIB with the dSPACE hardware and a power supply unit. Into the "servo module" were integrated the five servo amplifiers. The third module contains the power-source- and stepper-motor electronics.

To operate lifting magnets or electro-magnetic valves, an electronic unit with five switched voltage sources was newly developed (36 V/4 A). It was also integrated into the third module.

Figure 6 displays the TESLA hardware as a block diagram. You can also note the allocation of the individual electronic elements to the three aforementioned modules.

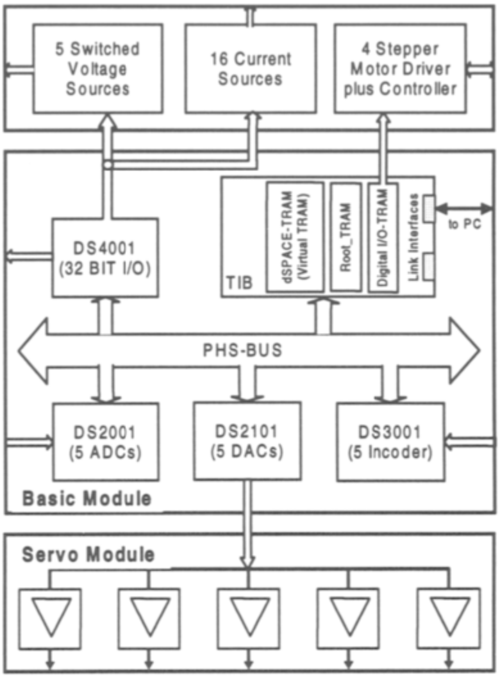

5 Switched Voltage Sources
16 Current Sources
4 Stepper Motor Driver plus Controller
DS4001 (32 BIT I/O)
TIB
dSPACE-TRAM (Virtual TRAM)
Root_TRAM
Digital I/O-TRAM
Link Interfaces
to PC
PHS-BUS
DS2001 (5 ADCs)
DS2101 (5 DACs)
DS3001 (5 Incoder)
Basic Module
Servo Module

Figure 6 – Block diagramm of the TESLA hardware

4. APPLICATION EXAMPLES

[1] already described the use of TESLA in the tests of a cash dispenser module and a stripe-card reader. We want to present as another example the use of the newly integrated modal-analysis function for an examination of the vibrations in a matrix printer.

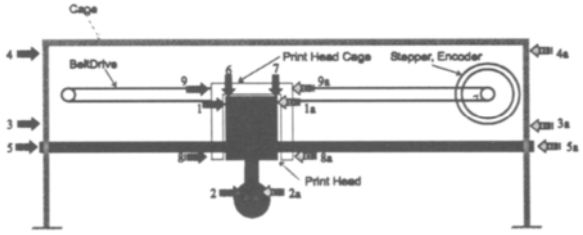

Cage
BeltDrive
Print Head Cage
Stepper, Encoder
Print Head
1
1a
2
2a
3
3a
4
4a
5
5a
6
7
8
8a
9
9a

Figure 7 – Schematic of a matrix printer

Figure 7 shows a schematic representation of those components of a matrix printer that are relevant for a vibration analysis. Additionally, the measurement points of the acceleration sensors (1...9a) for the modal analysis are marked. The print head resp. its caging are driven by a stepper-motor by means of a synchronous belt and deflection rollers. TESLA's stepper-motor electronics excited the system in the frequency range of about 15 Hz to 130 Hz. The measurement signals of the acceleration sensors can be measured by means of the ADC board. For the data measured in the time domain, the frequency spectra resp. the transfer functions could be determined with the help of the FFT analyser. In a next step, the modal-analysis tool can read the frequency-domain data and compute from them the modal-analysis data.

Figure 8 displays the results of the identification of two frequency-response functions (FRFs). The noisy signals are the measurement data, while the smooth graphs represent the identified FRFs. The FRFs in Figure 8 present two modes:

Figure 8 – Measured and identified FRFs of a matrix printer

The modal data established in the identification process are now taken to animate the oscillatory behaviour of the system. The animation depicts the oscillatory behaviour of the system in the case of an excitation in the modes and can give the construction engineer information on how to improve the system.

Another recent project of MLaP involves the modelling of an intermediate money storage. The intermediate money storage is part of an automatic teller machine. In this project, TESLA is used for identifying the parameters of the system, among other things.

SNI employs the TESLA system also for an experimental determination of stepper-motor sequences. SNI uses stepper motors both in automatic teller machines, passbook printers, and document printers. The motivation for developing a stepper-motor electronics of our own resulted from the fact that when we set about this task there was no suitable commercial electronics that met SNI's demands on power spectrum and flexibility.

At the university, the TESLA system is employed in the practical course accompanying the lecture on "The Application of Micro-electronics for Process Control" and in a class on modal analysis. The theoretical foundations, compiled during the lecture, concerning micro-processors, ADC-, DAC- and encoder-boards are deepened during the practical course.

The class on modal analysis deals with the aforementioned analysis of the vibrations of a matrix printer as an example of the practical application of modal analysis.

5. CONCLUSION AND OUTLOOK

The examples of TESLA's application prove the system to be a helpful tool for the analysis of mechatronic systems in precision mechanics. The system can also be employed efficiently for educational purposes in order to make obvious to the students the relevancy of the know-how acquired for the engineer's practical work.

Future work on TESLA will at first be focused on perfecting the operating control. Another main aspect of further developing TESLA will be the extension of the modal-analysis functions. We are going to implement more algorithms, make operating TESLA easier, and improve the animation. In the context of the HIL simulation possibilities inherent in TESLA, we plan to store the entire system, along with its signal-generation and signal-recording electronics, as a model in the computer.

6. REFERENCES

[1] Landwehr, M. and Robrecht, M.: *TESLA as Part of the Development Cycle for Mechatronic Systems.* 5th UK Mechatronics Forum International Conference, Guimarães, Portugal, 1996.

[2] Gaedtke, T.: *Entwicklung eines Transputer-Interface-Boards (TIB) als zentraler Baustein für eine industriell gefertigte Prozeßperipherie.* Final report on the ZIT project Modular Systems for Digital Simulation. MLaP, University of Paderborn, Germany, 1990.

[3] MSC Vertriebs GmbH: *MSC Transputer Products.* Karlsruhe, Germany, 1993.

[4] dSPACE GmbH: *DSP-CITpro Hardware (documentation and handbook),* dSPACE GmbH, Paderborn, Germany, 1991.

[5] LabWindows/CVI: *Visual Programming for Instrumentation.* National Instruments Corporation, Austin, Texas, 1996.

[6] Inmos Ltd. *ANSI C Toolset – Reference Manual.* UK, 1992.

[7] Nyenhuis, M.: *Entwurf und Realisierung eines Modalanalysewerkzeugs zur Schwingungsuntersuchung mechatronischer Systeme und Integration in eine bestehende Meßhardware und –software.* Diploma thesis, Mechatronics Laboratory Paderborn, University of Paderborn, 1997.

[8] Richert, J. and Hahn, M.: *DSS - DSL - DSC The Three Levels of a model Description Language for Mechatronic Systems.* ICMA '94 Mechatronics Spells Profitability, Tampere, Finnland, 1994.

[9] Homburg, C.: *Parallele Simulation mechatronischer Systeme auf der Basis von DSC.* 10. Workshop „Simulation verteilter Systeme und paralleler Prozesse" Dresden, Germany, 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

MECHATRONICS DESIGN OF AN INVERTED PENDULUM SYSTEM FOR ENGINEERING EDUCATION *

Celal S. Tüfekçi, Julian A. de Marchi, Kevin C. Craig [a], and Selahattin Özçelik [b]

[a]Department of Mechanical Engineering, Aeronautical Engineering, and Mechanics
Rensselaer Polytechnic Institute, Troy, NY, 12180, USA
E-mail: `tufekci@rpi.edu`
URL: `http://www.meche.rpi.edu/research/mechatronics/`

[b]Department of Mechanical and Industrial Engineering
Texas A&M University-Kingsville, Campus Box 191, Kingsville, TX, 78363, USA

ABSTRACT

This paper details the mechatronic design process as applied to the design of an inverted pendulum system, for educational use by undergraduate mechatronics students. Conception and development of the design are outlined with an emphasis on mechatronic design principles. Realisation of the design is made with simultaneous consideration towards constraints such as cost, reuse of existing materials, functionality, extensibility and educational merit. It represents a need for sensitivity towards these aforementioned issues, and also, how well the system represents the idealised dynamics presented in the classroom. The system design consists of "upgrading" an existing system, by replacing analogue sensors with digital ones, refitting some rotary bearings, and improving other design features wherever possible and appropriate. The resulting design is a linear-track inverted pendulum system with swing-arround capability, designed to familiarize students with control system design principles.

KEYWORDS: mechatronics, dynamic system investigation, synergism, integration, control system design, mechatronics design, modeling.

1. INTRODUCTION

Mechatronics is the *synergistic* combination of mechanical engineering, electronics, control systems and computers. The key element in mechatronics is the *integration* of these areas through the design process. The concept of synergism and integration through the design process is illustrated in Figure 1. The key to success in mechatronics system design is a balance between two skills:

- modeling (physical and mathematical), analysis (closed-form and numerical solution), and control system design (analogue and digital) of dynamic physical systems
- experimental validation of models and analysis (for computer simulation without experimental verification is at best questionable, and at worst useless) and understanding the key issues in hardware implementation of design

In order to design and build quality precision consumer products in a timely manner, the present-day mechanical engineer must be knowledgeable (both analytically and practically) in many different areas as indicated in Figure 1. The ability to design and implement analogue and digital control systems, with their associated analogue and digital sensors, actuators, and electronics, is an essential skill of every mechanical engineer, as almost everything today needs controls.

---

*This work is partially funded by Xerox, Inc.

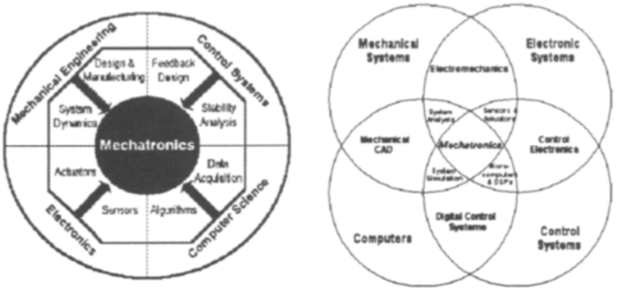

Mechanical Engineering
Control Systems
Electronics
Computer Science
Design & Manufacturing
Feedback Design
System Dynamics
Stability Analysis
Mechatronics
Actuators
Data Acquisition
Sensors
Algorithms
Mechanical Systems
Electronic Systems
Electromechanics
System Analysis
Sensors & Actuators
Mechanical CAD
Mechatronics
Control Electronics
System Simulation
Micro-computers & DSPs
Digital Control Systems
Computers
Control Systems

Figure 1. Mechatronics: Synergism and integration through design

Control design is not just for specialists anymore.

Figure 2 shows a diagram of the procedure for *dynamic system investigation* which emphasizes the balance between the aforementioned skills. The dynamic system investigation is a guide for the study of various mechatronic hardware systems. At Rensselaer, students perform a complete dynamic system investigation of mechatronic systems while they develop skills that will be indispensable as mechatronics design engineers in future years [2].

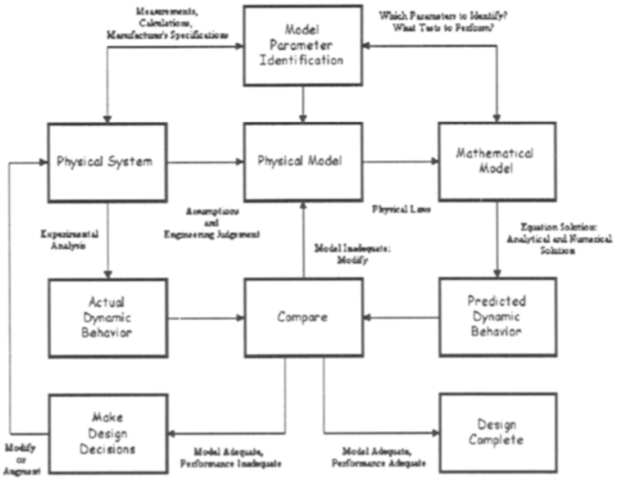

Measurements, Calculations, Manufacturer's Specifications
Model Parameter Identification
Which Parameters to Identify? What Tests to Perform?
Physical System
Physical Model
Mathematical Model
Assumptions and Engineering Judgement
Physical Laws
Experimental Analysis
Model Inadequate: Modify
Equation Solution: Analytical and Numerical Solution
Actual Dynamic Behavior
Compare
Predicted Dynamic Behavior
Make Design Decisions
Design Complete
Modify or Augment
Model Adequate, Performance Inadequate
Model Adequate, Performance Adequate

Figure 2. Dynamic system investigation

This paper focuses on the mechatronics des of an inverted pendulum system for engineering education. The presentation here evolves as an educational approach and reflects the coursework offered at Rensselaer.

The inverted pendulum system is a very popular system for controls education [1]. This system is chosen for mechatronic design since it covers almost every area of interest depicted in Figure 1. Some of the other systems used in mechatronics education at Rensselaer are the DC motor, stepper motor, thermal system, pneumatic system, and magnetic levitation system, to which a very similar design procedure is applied. However, it is worthed to mention that every system has its own characteristics.

The use of the dynamic system investigation procedure depicted in Figure 2 for the inverted pendulum system is discussed throughout the paper. The modeling of the inverted pendulum is explored in Section 2. Section 3 explains the details of the inverted pendulum system design for both electrical and mechanical aspects. Control system design and implementation is studied in Section 4 and identification for friction is discussed in Section 5.

2. SYSTEM MODELING

The real *physical system* is an inverted pendulum system which can swing around a pivot point attached on a cart. A force is applied to the cart to keep the cart position and the pendulum angular position, at the desired values, i.e. the cart centered and the stick balanced upwards. In order to keep this inherently unstable system under control, the required force is applied by an electric motor and the feedback is provided by optical encoders.

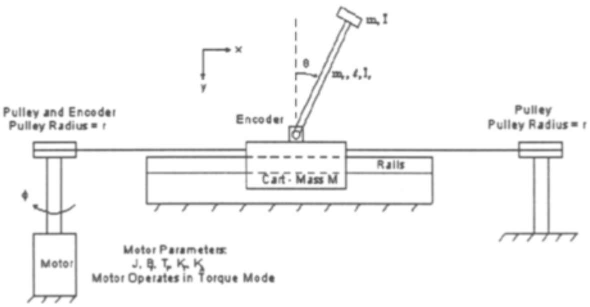
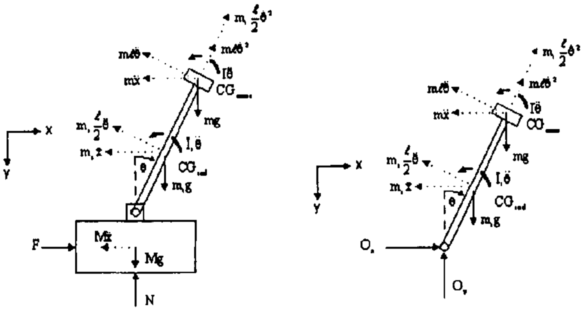

m, I
x
y
θ
m, ℓ, I
Pulley and Encoder Pulley Radius = r
Encoder
Pulley Pulley Radius = r
Rails
Cart - Mass M
ϕ
Motor
Motor Parameters: J, B, T, K, K
Motor Operates in Torque Mode

Figure 3. Inverted pendulum physical model

The ***physical model*** for the system is constructed using lumped physical elements, i.e., rigid rod, lumped mass, etc. After applying a set of simplifying assumptions along with use of engineering judgement, the physical model shown in Figure 3 results.

The *mathematical model* of the plant may be obtained by using several different methods, i.e., D'Alembert's, Newton's, Lagrange's, Kane's etc. One of the above methods is applied after drawing the free-body-diagrams depicted in Figure 4 and the equations (1) and (2) of motion are found which describe the non-linear plant dynamics.



Free Body Diagram: Cart & Pendulum

Free Body Diagram: Pendulum Alone

Figure 4. Inverted pendulum free-body-diagrams

$$\begin{aligned} F &= (M+m_r+m)\ddot{x} \\ &\quad +(m+\frac{m_r}{2})\ell\ddot{\theta}\cos\theta \\ &\quad -(m+\frac{m_r}{2})\ell\dot{\theta}^2\sin\theta \qquad (1) \\ 0 &= (\bar{I}_r+\bar{I}+m_r\frac{\ell^2}{4}+m\ell^2)\ddot{\theta} \\ &\quad +(m+\frac{m_r}{2})\ell\ddot{x}\cos\theta \\ &\quad -(m+\frac{m_r}{2})g\ell\sin\theta \qquad (2) \end{aligned}$$

We can now find linearised equations of motion about the operating points of $\bar{x}=0$, $\bar{\theta}=0$ in equations (3) and (4) by use of Taylor's series expansion.

The force appeared in equation (1) can be calculated using a torque generated by the electric motor. The dynamic relation for this transformation is given in equation (5) which may be substituted into equation (3) to obtain the relation between outputs, x, the cart position, θ, the pendulum angle, and inputs, V_{in}, the input voltage, T_f, the friction torque.

$$\begin{aligned} F &= (M+m_r+m)\ddot{x} \\ &\quad +(m+\frac{m_r}{2})\ell\ddot{\theta} \qquad (3) \\ 0 &= (\bar{I}_r+\bar{I}+m_r\frac{\ell^2}{4}+m\ell^2)\ddot{\theta} \\ &\quad +(m+\frac{m_r}{2})\ell\ddot{x} \\ &\quad -(m+\frac{m_r}{2})g\ell\theta \qquad (4) \\ F &= \frac{K_tK_aV_{in}}{r}-\frac{J\ddot{x}}{r^2} \\ &\quad -\frac{B_f\dot{x}}{r^2}-\frac{T_f[\mathrm{sign}(\dot{x})]}{r} \qquad (5) \end{aligned}$$

The description of model parameters along with numerical values for the inverted pendulum system is given in Table 1.

3. HARDWARE IMPLEMENTATION

The realisation of the design is made with simultaneous consideration towards constraints such as cost, reuse of existing materials, functionality, extensibility, and educational merit. The inverted pendulum system is inherited from previous years and upgraded to its current state by taking above issues into account. The analogue potentiometers to measure positions are replaced with optical encoders, which are noise-free. The linear track friction is critically reduced and the transmission resolution is improved by replacing a smaller diameter drive wheel. Swing-arround capability is added to the inverted pendulum system to allow design a swing-up controller for self erection. The necessary parts are machined or purchased within a budget, or reused where possible, and assembled to the design specifications.

The inverted pendulum system interface is designed with consideration to the mechanical design. The interface includes selection of hardware actuators, sensors and signal processing, software for interpretation of the sensor signals, and their synthesis to form a closed-loop control signal. The inverted pendulum system and its peripherals form a compound interface depicted in Figure 5.

Table 1
Inverted pendulum parameters

SYMBOL	PARAMETER DESCRIPTION	VALUE	UNIT
g	gravitational acceleration	9.81	m/s
M	mass of the cart	0.965	kg
ℓ	length of the pendulum	0.636	m
m_r	mass of the pendulum	0.0576	kg
m	mass of the end weight	0.116	kg
$\bar{I}_r$	moment of inertia of pendulum	0.0019	kgm^2
$\bar{I}$	moment of inertia of end mass	1.3127×10^{-5}	kgm^2
J	motor inertia	1.8367×10^{-4}	kgm^2
K_a	servo amplifier gain	1.6	A/V
K_t	motor torque constant	0.1013	Nm/A
B_f	motor viscous friction	1.33×10^{-4}	Nms/rad
r	radius of the pulley wheel	0.01135	m

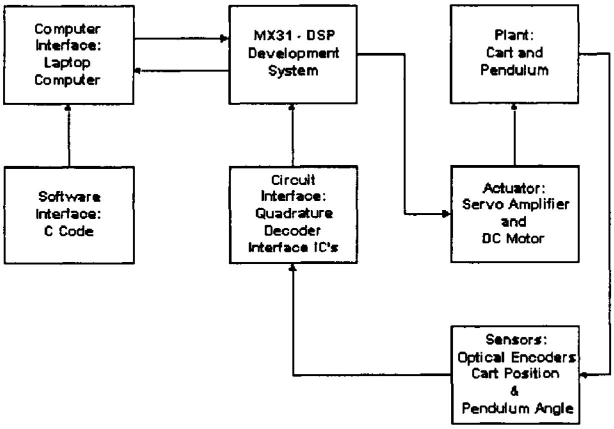

Computer Interface: Laptop Computer
MX31 - DSP Development System
Plant: Cart and Pendulum
Software Interface: C Code
Circuit Interface: Quadrature Decoder Interface IC's
Actuator: Servo Amplifier and DC Motor
Sensors: Optical Encoders Cart Position & Pendulum Angle

Figure 5. Inverted pendulum and peripherals interface

The transducer placement, precision, and range are integral to both the mechanical and electronics design. Calculations are made to specify the sensing and actuating requirements for the system, and mechanical and electrical components are selected which will work together in a complete design before anything is actually built. The overall system is identified to verify the idealisability of its dynamic behavior.

The complete mechatronic system design thus includes design concept generation, mechanical dynamics analysis, dynamics simulation, component selection and fabrication, electronic hardware and transducer selection and interfacing, circuit design and wiring, software design, system dynamics identification and model verification, and finally controller design and implementation. Each of these steps are highlighted within the scope of the inverted pendulum system, and how mechatronic design principles are key to achieving success on the *first* iteration of the design process.

4. CONTROL SYSTEM DESIGN AND IMPLEMENTATION

The control system architectures are developed for the generic system, and sample controllers are designed for the nominal plant parameters given in Table 1. This is done concurrently with the system development, and once the system identification is performed, actual parameter values may be used to refine the controllers.

The basic approach in the control system design for the inverted pendulum system is to first design the controller in a conventional manner and then discritise it so that the resulting controller can be coded on the computer easily. The C programming language and compilers are used to construct and compile the source code to generate machine code. Then, the object code is dowloaded to an embedded DSP which is the controller running at $f_s = 500\ Hz$.

The control objective for the inverted pendulum system is to keep the pendulum upright in the presence of gravity and unexpected force disturbances applied to the cart or to the pendulum, and to keep the cart centered on the track. In other words, given any initial conditions caused by disturbances, the pendulum can be brought back to the vertical position and also the cart can be brought back to the reference position quickly (e.g., settling time, $t_s \approx 2$ *sec*) with reasonable damping (e.g., $\zeta \approx 0.5$).

The classical, state-space, and fuzzy logic controllers are designed in SIMULINK^R and implemented on the actual inverted pendulum system. The flowchart shown in Figure 6 explains the general control algorithm implementation in the C programming language.

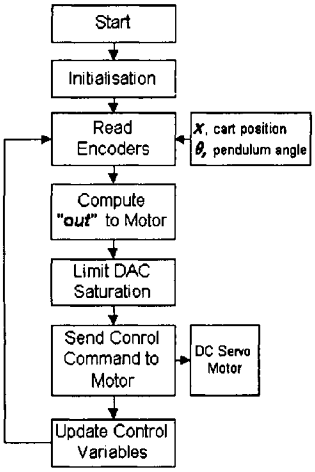

Start
Initialisation
Read Encoders
X, cart position
θ, pendulum angle
Compute "out" to Motor
Limit DAC Saturation
Send Conrol Command to Motor
DC Servo Motor
Update Control Variables

Figure 6. Control algorithm

In the classical control design, first a continuous controller designed for the stick balance, which is called the inner-loop controller as seen in Figure 7. Then, the outer-loop controller designed to bring the cart to the center of the track. Root-locus design techniques are used in the design process and the controllers given in equations (6) are implemented on the DSP.

$$\begin{aligned} G_1(s) &= +0.6\frac{s+1}{s+15} \\ G_2(s) &= -23.5\frac{(s+1.5)(s+3)}{s(s+20)} \end{aligned} \qquad (6)$$

The controllers given in equation (6) are discritised at a sampling rate of $T_s = 0.002$ *s* using Tustin's method (bilinear transformation [3]) and the resulting difference equations are implemented in the C source code.

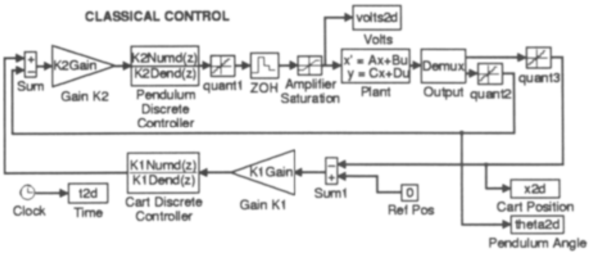

CLASSICAL CONTROL
volts2d
Volts
Sum
K2Gain
Gain K2
K2Numd(z)
K2Dend(z)
Pendulum Discrete Controller
quant1
ZOH
Amplifier Saturation
x' = Ax+Bu
y = Cx+Du
Plant
Demux
Output
quant2
quant3
K1Numd(z)
K1Dend(z)
Cart Discrete Controller
K1Gain
Gain K1
Sum1
0
Ref Pos
x2d
Cart Position
theta2d
Pendulum Angle
t2d
Clock
Time

Figure 7. Classical controller

The state-space controller is designed using the linear quadratic state feedback regulator (LQR) method. The linear velocity for the cart and the angular velocity for the pendulum are estimated using a finite difference approximation. The LQR control system simulation is shown in Figure 8 and the following gains given in equation (7) are obtained.

$$K = [-3.1623 \; -3.2488 \; -15.8530 \; -3.9252] \quad (7)$$

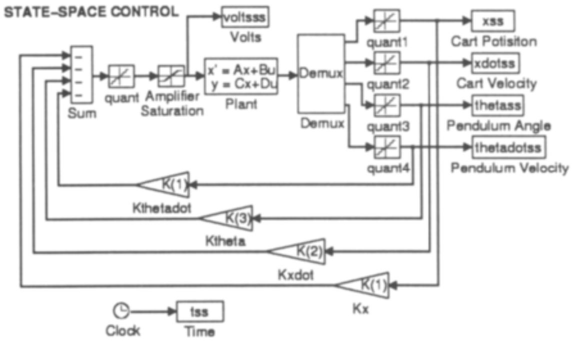

STATE-SPACE CONTROL
voltsss
Volts
Sum
quant
Amplifier Saturation
x' = Ax+Bu
y = Cx+Du
Plant
Demux
Demux
quant1
quant2
quant3
quant4
xss
Cart Potision
xdotss
Cart Velocity
thetass
Pendulum Angle
thetadotss
Pendulum Velocity
K(1)
Kthetadot
K(3)
Ktheta
K(2)
Kxdot
K(1)
Kx
Clock
tss
Time

Figure 8. State-space controller

The fuzzy logic controller is designed using the *Fuzzy Logic Toolbox* for MATLAB^R. There are 4 pair of membership functions associated with

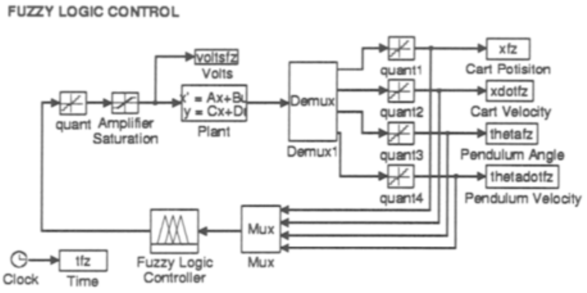

FUZZY LOGIC CONTROL
voltsfz
Volts
x' = Ax+Bu
y = Cx+Du
Plant
quant
Amplifier Saturation
Demux
Demux1
quant1
quant2
quant3
quant4
xfz
Cart Potisiton
xdotfz
Cart Velocity
thetafz
Pendulum Angle
thetadotfz
Pendulum Velocity
Mux
Mux
Fuzzy Logic Controller
Clock
tfz
Time

Figure 9. Fuzzy logic controller

each state of the inverted pendulum system. The simulation block diagram is given in Figure 9.

The simulation results are presented in Figure 10 for each of the above three simulations. The experimental results obtained from the state-space controller implementation are shown in Figure 11. We add the dry friction compensation to avoid the dynamic lags and the stick-slip phenomenon. The first of the three plots in Figures 10 and 11 represents the control effort, V_{in} or **out** in *volts*, the second shows the cart position, x in *metres*, and the third depicts the pendulum angle, θ in *radians*. The reason that the experimental results shown in Figure 11 have so many jitter is due to the friction between the cart and the linear track.

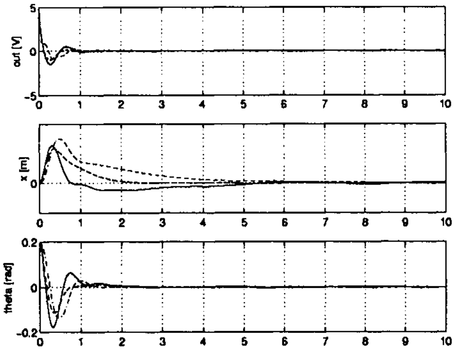


Figure 10. Simulation results (– classical, -- LQR, -.- fuzzy)

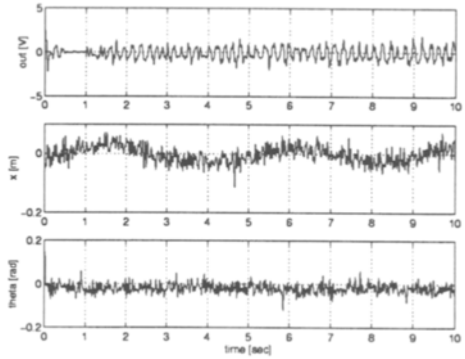


Figure 11. Experimental results for LQR

5. FRICTION IDENTIFICATION

The idealised inverted pendulum system does not exhibit any friction characteristics in the sense that we want the student to conceptually understand the dynamic behaviour of the system and to model and control that behaviour. Including frictional effects complicates the dynamic analysis. Wet (viscous) friction is the only kind which can be linearly modeled, however in the actual system the overall friction acts predominantly like dry (Coulomb) friction, since the bearing lubricant is only a very thin layer of very thin oil. The controller design would therefore require the analysis of a nonlinear system, which is beyond the scope of the coursework. However, the fact that the actual system exhibits significant frictional effects demands that the issue be addressed.

The simplest possible friction model was therefore adopted, which is the traditional textbook model of wet+dry friction as shown in Figure 12.

Because the friction is piecewise continuous, the system is linear as long as the sign of the velocity does not change. This means an analytical solution for the dynamic system can still be obtained, where the solution spans regions of either positive or negative velocity, and the final conditions for one region become the initial conditions for the following region across the sign change. When the velocity is zero and the applied force is less than the static force, the system is simply static until the necessary force threshold is reached by the motor.

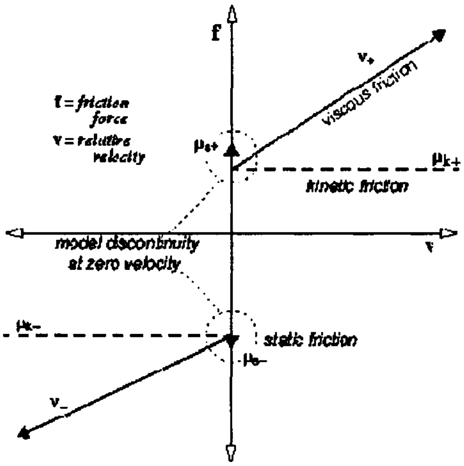


Figure 12. Friction model

To simplify the friction modeling for simulation and controller design purposes, the positive and negative dry friction values are assumed to be equal and opposite, and any discrepancy is treated as a modeling error which is compensated for via some integral control action.

To identify the actual friction in the system, the motor is ramped up slowly from zero force, and the moment the cart begins to move, the applied force is frozen and recorded. This force is thus equivalent to the actual static friction force. After the cart reaches a steady speed the force is next reduced from just above the static friction force value until the cart stops. This is thus the actual kinetic friction force. The viscous friction is estimated by applying a step force to the motor and observing the rise time of the cart velocity.

6. CONCLUSIONS

The mechatronics design of an inverted pendulum system for engineering education is achieved. After applying the dynamic system investigation process on this system, it is well understood and a compact design result is accomplished. The modeling, analysis, and control system design are considered along with experimental validation and the important key issues in hardware implementation. Since we take an educational approach in engineering, the control system design is started from a very conventional way, i.e., classical control design, and the other types of controllers are demonstrated without going into too much details of the theory.

The friction has very important effects on the modeling and control system design such that it causes a drift of the cart to one way in the classical control implementation. For this reason, the static friction is identifed and its effect is reduced on the system by placing a simple friction compensation.

In the future, it is planned to add a swing-up controller so that the pendulum can be erected by making the system unstable in some acceptable range. Then, one of the studied controllers may be switched when the stick is at the upright position and the control objectives addressed in this paper may be accomplished again. Furthermore, some other types of controllers may be designed to make students more familiar with the principles of control system design.

REFERENCES

1. Shaian, B. and Hassul, M., *Control System Design using* MATLAB, Prentice-Hall, 1993.
2. Craig, K.C., "Mechatronics in University and Professional Education: Is there anything really new here?", The Winter Annual Meeting of ASME, to appear, Anaheim, CA, November 1998.
3. Franklin, G.F., Powell, J.D., Workman, M.L., *Digital Control of Dynamic Systems*, 2^{nd} Edition, Addison Wesley, 1994.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Enhancing a Mechanical Engineering Curriculum with a Mechatronics Theme

David G. Alciatore and Michael B. Histand

Department of Mechanical Engineering
Colorado State University
Fort Collins, CO 80523
USA

This paper presents a stepwise restructuring of a mechanical engineering curriculum with a mechatronics theme. The result can be a curriculum with contemporary emphasis and enhanced content realized by an improved sequencing of modeling and analysis, computing, electrical circuits and machines, measurements and instrumentation, microprocessors, and design. Mechatronics will provide the focus for problems, exercises, and design experiences. The proposed curriculum is envisioned as a four-step evolutionary process that allows incremental changes coordinated with faculty development, laboratory design, and course resources.

1. INTRODUCTION

Although an emphasis on mechatronics should not change the fundamental content of a mechanical engineering curriculum, it does provide the opportunity to contemporize some subject matter, enhance content, and improve sequencing of modeling and analysis, computing, electrical circuits and machines, measurements and instrumentation, microprocessors, and design. Furthermore, a central curricular theme of mechatronics provides a better focus to connect the intellectual growth that occurs in undergraduates as they proceed through the four years of an ABET-accredited program. Disjoint, compartmentalized curricular structures that separate curricular topics into unconnected pedagogical packages make it difficult for some students to comprehend and embrace the concept of mechanical engineering system design. Today's students want to see how the foundation studies are germane to the practice of design engineering, and mechatronics provides an excellent focus to inspire and motivate students. The word mechatronics even symbolizes the holistic approach to learning that enables us to provide this focus. A mechatronics theme can help make the mechanical engineering curriculum more stimulating, meaningful, fun, and attractive, and can better prepare students to pursue careers that are characterized by a high degree of innovation and leadership.

It is difficult to dramatically redesign a traditional mechanical engineering curriculum with a mechatronics theme in one revolutionary step. The difficulty is due to the time required to gain consensus among the faculty and administrators, the need to train and/or hire faculty, and the need to develop laboratories that are necessary to assure that teaching and learning meet the new goals of the proposed structure.

2. CURRICULUM EVOLUTION

As illustrated in the figures below, we present a model for a stepwise change in the mechanical engineering curriculum expanding around the theme of mechatronics. We propose as one approach, a four-stage evolution that will recast some courses, establish better links between courses, and define new courses.

Figure 1 shows the typical courses in a mechanical engineering curriculum germane to the field of mechatronics. The course titles come from a past Colorado State University curriculum, but most universities have similar courses. In this figure and the figures that follow, arrows indicate sequencing or prerequisites and boxes indicate courses with common subject matter and close linkages. In subsequent figures, a shaded course block indicates that the course is the focus of change or restructuring in the current evolutionary step.

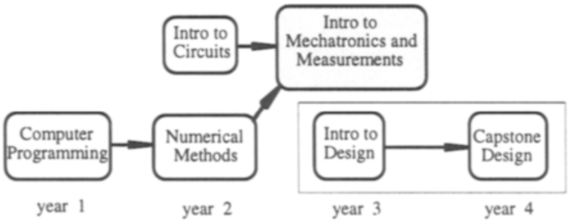

Intro to Circuits
Intro to Mechatronics and Measurements
Computer Programming
Numerical Methods
Intro to Design
Capstone Design
year 1
year 2
year 3
year 4

Figure 1 Typical mechatronics theme foundation courses

Figure 2 illustrates the first step in the evolution to create an introductory mechatronics course that emphasizes system response, measurement fundamentals, digital and analog electronics, sensors, actuators, and basic mechatronic system control architectures. We have previously described early versions of this course at Colorado State University [1, 2]. This course could serve as a replacement for a traditional measurements, instrumentation, and/or modeling course, or it could be a new course. In either case, faculty development in modern electronics and applications may be required to ensure the course's success. Another option may be team teaching the course with colleagues from an electrical engineering department.

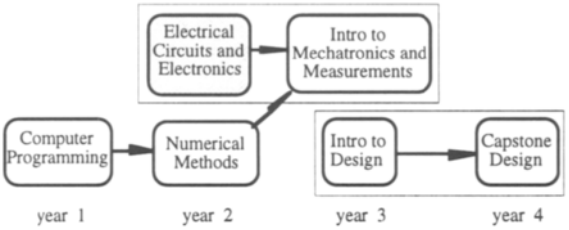

Electrical Circuits and Electronics
Intro to Mechatronics and Measurements
Computer Programming
Numerical Methods
Intro to Design
Capstone Design
year 1
year 2
year 3
year 4

Figure 2 First step in curriculum evolution

Figure 3 illustrates the second step of the evolution to transform a traditional course in electric circuits to provide a better foundation for the mechatronics course. The transformed course can include Kirchoff's Laws, ac and dc circuits and sources, computer-aided circuit design and analysis, basic electrical measurements, simple op amp circuits, basic digital circuits, dc motors, optoelectronics, and electrical safety. If possible, the subject matter should be coupled with some laboratory experiences that require building circuits and using the oscilloscope, digital multimeter, and power supplies. Throughout the evolution of the mechatronics theme courses, effort must be devoted to creating laboratory exercises and design projects that are stimulating, somewhat open-ended, and appropriate to previous experience and future context. Although typical "Intro to Circuits" courses are taught by electrical engineering faculty, we recommend that the proposed "Electrical Circuits and Electronics" course be taught within mechanical engineering to help provide students better context. We have recently published a textbook [3, 4] that has helped us develop the first and second steps of the evolutionary process.

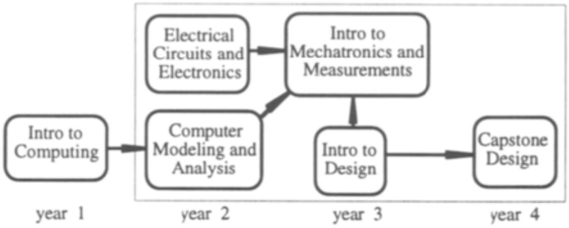

Electrical Circuits and Electronics
Intro to Mechatronics and Measurements
Intro to Computing
Computer Modeling and Analysis
Intro to Design
Capstone Design
year 1
year 2
year 3
year 4

Figure 3 Second step in curriculum evolution

Figure 4 illustrates the third step in the evolution to transform a traditional computer applications course (e.g., "Numerical Methods") to better dovetail into the mechatronics theme by infusing topics related to system modeling and analysis, and an introduction to microcontrollers and other control architectures (e.g., single board computers and PLCs). The computing learned in the course ("Computer Modeling and Analysis") would be designed in part to provide foundations for the mechatronics course: digital representations and number systems, simple interfacing, and microcomputer architectures. Interfacing, programming, feedback control, and more sophisticated exercises and design problems can also be infused into the course. The box shown in Figure 4 contains three 3-credit courses, but we feel that they could be integrated well enough to permit packaging the material into two 4-credit courses resulting in a reduction of total credits. We also propose changing the "Intro to Computing" course away from traditional programming towards application-based computer solutions. This transformation is described elsewhere [5].

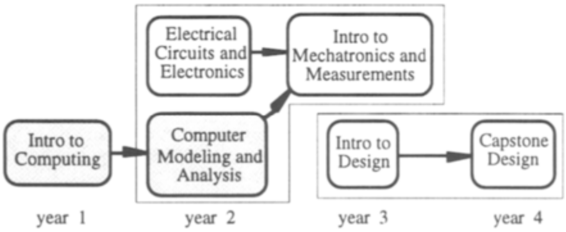

Electrical Circuits and Electronics
Intro to Mechatronics and Measurements
Intro to Computing
Computer Modeling and Analysis
Intro to Design
Capstone Design
year 1
year 2
year 3
year 4

Figure 4 Third step in curriculum evolution

Figure 5 illustrates the fourth step in the evolution of the curriculum to include an introductory design course as a prerequisite to the mechatronics course. The design methodology and project management skills taught in the design course, in addition to preparing students for their capstone experience, can also be used to help ensure the success of project activity in the mechatronics course. With this prerequisite, students will be better prepared to focus on the formal design process, planning and scheduling, optimization, computer aided design and other elements necessary to facilitate their design experiences in mechatronics and the capstone design course. As the large box in Figure 5 implies by including "Capstone Design," the mechatronics theme and improved integration also enables more sophisticated mechatronics capstone design projects.

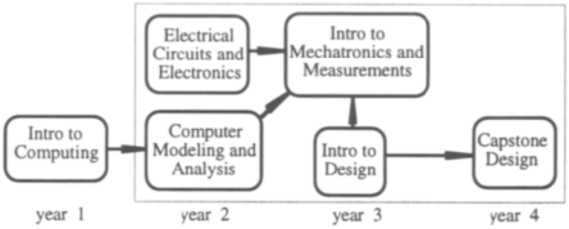

Electrical Circuits and Electronics
Intro to Mechatronics and Measurements
Intro to Computing
Computer Modeling and Analysis
Intro to Design
Capstone Design
year 1
year 2
year 3
year 4

Figure 5 Fourth step in curriculum evolution

Throughout the stepwise evolution, we propose replacing or dramatically changing some traditional courses in electrical circuits, computing, instrumentation and measurements, and design methodology to better integrate them around a mechatronics theme. A result is improved course efficiency created by presenting the various topics in a more integrated way with more unified context, which can also result in enhanced student motivation and learning.

3. CONCLUSIONS

In summary, we presented a stepwise enhancement of a mechanical engineering curriculum that responds to current issues in mechatronics, providing students with a major educational theme that will keep them invested in and excited about their learning. The proposed changes result in a curriculum with enhanced content and improved sequencing of foundation mechatronics topics. Different universities may have different courses and emphases and may elect to skip some or all of the incremental changes, although we did not believe that this was manageable for our program. Also, for us, the goal was not to create a novel "mechatronics program"; but rather, to give our mechanical engineering program better focus. The changes we propose may be integrated in whole or in part by departments based on their strengths and resources.

For descriptions of other programs that have substantial mechatronics components in their curricula, readers are referred the Proceedings of the Workshop on Mechatronics Education [6] and the Internet [7].

REFERENCES

1. Alciatore, "A Modern View of 'FORTRAN' and Drafting for Mechanical Engineering Freshmen," Proceedings of the ASEE Rocky Mountain Region Annual Conference, Denver, April, 1998.
2. Alciatore and Histand, "Mechatronics and Measurement Systems Course at Colorado State University," Proceedings of the Workshop on Mechatronics Education, sponsored by the Synthesis Coalition of the National Science Foundation, pp. 7-11, Stanford, CA, July, 1994.
3. Alciatore and Histand, "Mechatronics at Colorado State University," Mechatronics, v. 5, n. 7, pp. 799-810, Pergamon Press, 1995.
4. Histand and Alciatore, Introduction to Mechatronics and Measurement Systems, McGraw-Hill, New York, 1998.
5. http://www.engr.colostate.edu/~dga/mechatronics.html, "Mechatronics and Measurement Systems."
6. Proceedings of the Workshop on Mechatronics Education, sponsored by the Synthesis Coalition of the National Science Foundation, Stanford, CA, July, 1994.
7. http://www.mechatronik.uni-linz.ac.at/international/, "Mechatronics Around the World."

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

FUZZY CONTROL APPLICATIONS DEVELOPMENT SYSTEM

Jacobo Alvarez[a], Antonio Manzanedo[b].

[a]Department of Electronic Engineering, University of Vigo, Apartado oficial, 36200-VIGO, Spain.
[b]Department of Systems and Automation Engineering, University of Vigo, Apartado Oficial, 36200-VIGO, Spain.

One of the major challenges of an educator is to introduce new technologies in such a practical way that undergraduates can compare them with existing technologies. To achieve this, most of the time it is necessary to develop an special equipment because there is no such commercially available system, or it is expensive or inadequate.

The goal of this work is to design and build an equipment that permits the development and implementation of fuzzy control applications to teach this new technique and compare it with classical control.

1. HARDWARE

The first decision was to use a general purpose microprocessor instead of a microcontroller to choose independently the peripherals such as A/D and D/A converters to achieve the required characteristics. The decision is very difficult since there are a lot of candidates, but in order to reduce the cost of the equipment, the Intel 8088 was chosen. This microprocessor has enough capabilities to be the core of our system, but the key reason to choose it was the possibility to recycle all the components of old personal computers (PCs) that are still stored in many warehouses of our University.

The system RAM memory is composed of two banks and can be configured between 64 Kbytes and 512 Kbytes depending on the RAM chips availability and our needs. We choose to use the same dynamic RAM chips that came with personal computers, so the 8237A, DMA controller, was included to refresh the data in memory.

The system EPROM memory is composed of two sockets for 2764 (8Kx8 each), one intended to store the boot and fuzzy monitor routines and the other prepared to store the application code in stand-alone operation.

Communications are implemented via RS-232 and RS-485 standards.

The interface with the analog world was achieved through an A/D converter MAX180 and two D/A converters MAX526, both from Maxims. The first one provides 8 multiplexed input channels with 12 bit resolution and 10 us. conversion time. The D/A converters have four 12 bit converters each and 3 us. settling time. They are fast enough for the intended applications in Mechatronics.

There are, of course, some other complementary circuits like the 8253 programmable timer, buffers, registers and 'glue' logic.

2. SOFTWARE

The boot and monitor routines that are located in the first EPROM of the equipment were developed in assembler language so they are optimized for both speed and code size. The software that runs on the personal computer, intended to communicate with the equipment, was written in Pascal and, at this time, the interface is DOS like, not very user friendly.

The definition of the inputs, membership functions, rules, etc. is done through an ASCII file with a format defined below.

3. OPERATION MODES

One of the objectives of this project was to design an equipment with the maximum flexibility so it could be used, not only to develop fuzzy control applications (the main goal), but to be used as a general data acquisition system or to implement non-fuzzy controllers. With this in mind, several working modes were developed, some computer attached and some stand-alone.

- **Mode 0**: local test. In stand-alone operation, the equipment reads all analog and digital input channels and writes the values to the corresponding analog and digital output channels. This mode is intended for channel test only.

- **Mode 1**: memory read. In this mode, communication between a PC and the equipment is established and we can read any code or data stored in the RAM or EPROM modules. It is intended to test the communications link and the memory blocks.
- **Mode 2**: slave I/O operation. In this mode, we can use the equipment as a data acquisition system, having access to all inputs and outputs from another equipment. Working in this mode, we can implement in a computer any type of control algorithms over the external inputs and transmit the new outputs to the equipment.
- **Mode 3**: program load. In this mode, it is possible to load a new monitor program in the equipment RAM memory, so we can use this new program instead of the monitor program stored in the equipment EPROMs.
- **Mode 4**: stand-alone operation. In this mode, the equipment transfers the control to the monitor program stored in the equipment EPROMs. In that way, we can implement any kind of stand-alone controller.
- **Mode 5**: slave fuzzy operation. This is one of the most important modes, that allows the equipment configuration as a fuzzy control system via the RS-232 port. The computer sends the information about the membership functions of each input and output and the rules.
- **Mode 6**: not defined. For user purposes.
- **Mode 7**: stand-alone fuzzy operation. Instead of receiving the configuration through the RS-232 port, like in mode 5, this data is stored in the equipment EPROMs, so the system can work alone.

4. FEATURES

- The equipment can implement two or more independent controllers simultaneously, restricted to general characteristics.
- The analog inputs are fuse protected.
- The maximum number of inputs are 8, either analog or digital.
- The resolution of the analog inputs is 10 bits.
- Each input can have up to 8 fuzzy sets or regions.
- Each region must be defined by a membership function of any polygonal shape.
- The degree of membership can take values between 0 and 127.
- Each rule can have up to 256 complex antecedents. A complex antecedent is one formed by two or more simple ORed antecedents. A simple antecedent is formed by up to 8 ANDed antecedents where each antecedent (pair input-region) has to be related to a different input variable. An individual rule is one with simple antecedents. A multiple rule is one with complex antecedents.
- Each application permits a maximum of 256 individual rules or a combination of individual and multiple rules.
- The defuzzyfication method is the centroid one.

5. SOLUTIONS

To obtain the value of each output of a fuzzy controller in a given instant, implies a heavy computation load. This work has to be done in a short time to achieve the performance of a real-time controller, even in slow systems, such as motors, etc..

Another possibility is to tabulate all the values that can be obtained from this calculation. This is possible because the inputs can only have a finite number of values, depending on the number of resolution bits of the A/D converter, and there are a finite number of membership functions, previously defined for each application, and a finite number of rules. However, it requires a high amount of memory since the table size depends exponentially on the number of inputs, the number of resolution bits and lineally with the number of outputs.

For each output it is necessary a table which size will be $(2^{\text{n° of bits}})^{\text{n° of inputs}}$.

To reach a compromise between computation power and memory size needed, we decided to use a semi-tabular method, that is composed by the following steps (figure 1).

1) With the inputs, their regions and each region membership function defined, we build a **membership table** that, for every possible value of the input, gives the degree of membership to each region (1 byte). So, the fuzzyfication method consists only in accessing this table. The table size is given by the maximum number of inputs (8), the maximum number of regions per input (8) and the possible values of each input (1024 = 10 resolution bits). So, the size is 8x8x1024 = 64 Kbytes.

2) With the outputs, their regions and each region membership function defined, we build an **area table** that, for every possible value of the output, gives the total area below that value in the membership function and the momentum related to the origin. The table size

is given by the maximum number of outputs (8), the maximum number of regions per output (8) and the possible values of the degree of membership (128) and the bytes needed to store the area and momentum (8). So, the size is 8x8x128x8 = 64 Kbytes.

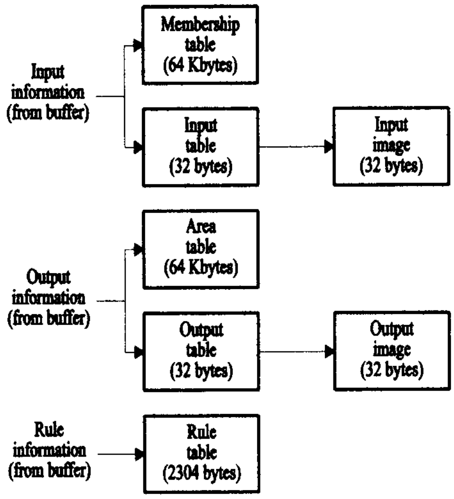

Input information (from buffer)
Membership table (64 Kbytes)
Input table (32 bytes)
Input image (32 bytes)
Output information (from buffer)
Area table (64 Kbytes)
Output table (32 bytes)
Output image (32 bytes)
Rule information (from buffer)
Rule table (2304 bytes)

Figure 1. Table calculation.

3) With the rules, we build a **rule table**. The table size is given by the maximum number of individual rules (256) and the maximum number of antecedents (8) that form a simple antecedent, the number of bytes needed to represent the region involved in each antecedent (1) and the region of the consequent of the rule (1). So, the total size is 256x(8x1+1x1) = 2304 bytes.

4) With the definition of the inputs, we build an **input table** that gives information about the input type (analog, digital, unipolar, bipolar), the number of regions described, etc.. This table is 32 bytes wide. When the application is running we build another table called **input image**, that contains the input type and the present value of each input and it is 32 bytes wide, too. In the same way, we build the **output table**, auxiliary for calculating the value of the outputs while the system is running, and the **output image**, each 32 bytes wide.

All these tables, except the input and output images, do not change their contents once defined for a specific application. When we initiate the process, that is, the fuzzy control of the target system, first the table values are calculated and memorized in the equipment RAM and then, each processing cycle is composed by the following steps (figure 2).

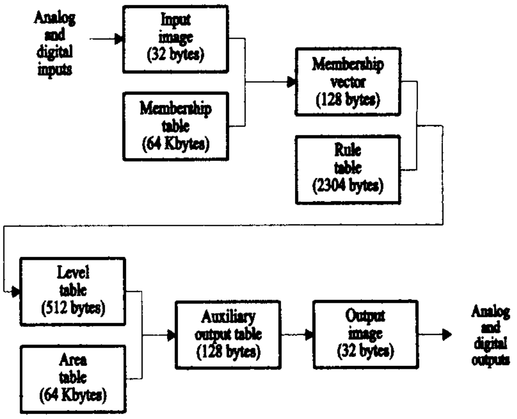

Analog and digital inputs
Input image (32 bytes)
Membership table (64 Kbytes)
Membership vector (128 bytes)
Rule table (2304 bytes)
Level table (512 bytes)
Area table (64 Kbytes)
Auxiliary output table (128 bytes)
Output image (32 bytes)
Analog and digital outputs

Figure 2. Cycle execution

1) The values for the inputs are obtained, storing them in the input image.

2) The membership table is used to obtain the degree of membership to each region for every input.

3) With these degrees of membership (x) and their complementaries (127-x) we build a **membership vector**, 128 bytes wide.

4) With this vector, the rules are applied through the rule table, obtaining the consequent (pair output-region) and calculating the value of the degree of membership for this rule. This information is stored in the **level table**.

5) The defuzzyfication is achieved by adding the areas and momentums related to the fired rules. The values for each output are calculated using the information of the level and area tables and are stored in the **auxiliary output table**, 128 bytes wide.

6) The final value of each output is obtained dividing total momentum by total area and storing the results in the output image. These values are written to the outputs in the end of the cycle.

Although integer arithmetics is used to simplify the calculation, the maximum errors do not reach 3 units in 1024, this is, 13 mV. in a range of 0-5 V.

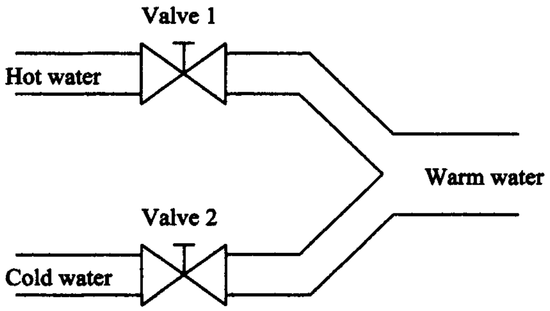


Figure 3. Home plumbing.

6. PRACTICAL RESULTS

Using a considerable number of rules, 64, and 4 inputs with an average number of regions per input of 4, the total cycle time is below 14 ms., that is, about 70 cycles per second, enough for many typical mechatronics applications.

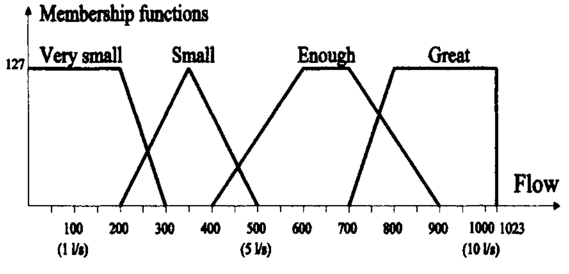


Figure 4. Membership functions for Water Flow.

7. TUTORIAL EXAMPLE

We have chosen a modern home application, water temperature and flow control to show the powerness and simplicity of this equipment. As you can see in figure 3, there are two valves to control the flow of cold and hot water and two sensors to measure the temperature and the flow of the resulting mixture.

The memberships functions that we have defined for the flow (F), the temperature (T) and the valve aperture (V) are in figures 4, 5 and 6. The rules in table 1 represent the valve aperture (output) depending on the values of flow and temperature (inputs). The value "nothing" means that the last value is not changed. With these rules, the processor

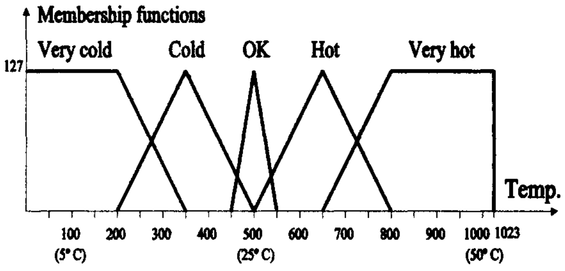


Figure 5. Membership functions for Water Temperature.

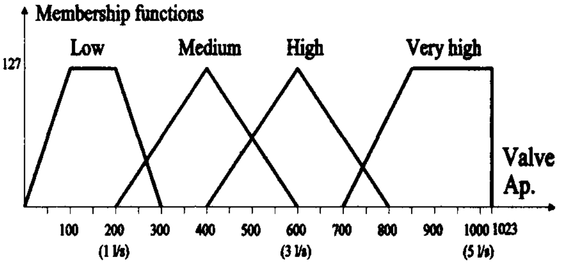


Figure 6. Membership functions for Valve Aperture.

Table 1
Valves aperture (fired rules)

	Flow			
	Very small	Small	Enough	Great
Temperature				
Very cold	Cold nothing Hot very high	Cold nothing Hot very high	Cold medium Hot very high	Cold low Hot high
Cold	Cold nothing Hot high	Cold nothing Hot high	Cold medium Hot high	Cold medium Hot nothing
OK	Cold very high Hot very high	Cold high Hot high	Cold nothing Hot nothing	Cold medium Hot medium
Hot	Cold high Hot nothing	Cold high Hot nothing	Cold high Hot medium	Cold nothing Hot medium
Very hot	Cold very high Hot nothing	Cold very high Hot nothing	Cold very high Hot medium	Cold high Hot low

executes 150 cycles/s, so the response time of the controller is optimal and negligible to the potential user.

We have tried some other applications like direct current motor velocity control and simulation of the Sugeno's mobile and the results are very satisfactory in precission and speed.

8. CONCLUSIONS

To summarize this work, we can stand out the following characteristics:
- Compact, strong and easy to use fuzzy application development equipment.
- Reduced hardware costs since many components are recovered from old PCs.
- The computational method uses tables as well as integer calculation, achieving a reasonably speed with a low amount of memory.
- Flexibility is one of the keys, so this equipment can be used as a data acquisition system or as any other (non-fuzzy) control application development system.
- It works on both slave (computer attached) and stand-alone modes.
- The code is written in assembler so it is optimized.

In figures 7, 8 and 9 you can see the fuzzy application development equipment.

9. FUTURE DEVELOPMENTS

- Development of Windows-based interface.
- Reduce the board size, using programmable circuits such as FPGAs.
- Increase the computation speed, substituting the 8088 for another faster microprocessor.

10. BIBLIOGRAPHY

[BRUBAKER 92] BRUBAKER, David I., SHEERER, Cedric, "Fuzzy-logic system solves control problem", *EDN*, June 18, 1992.

[CONNER 93] CONNER, Doug, "Designing a fuzzy-logic control system", *EDN*, March 31, 1993.

[CUBILOT 95] Cubilot Rodríguez, Marta, "Aplicación de la lógica borrosa al control de velocidad de un motor de corriente continua", Proyecto Fin de Carrera, Universidad de Vigo, Vigo, 1995.

[LEGG 93] LEGG, Gary, "Microcontrollers embrace fuzzy logic", *EDN*, September 16, 1993.

[MANZANEDO 97] Manzanedo García, Antonio, "Equipo para el desarrollo de aplicaciones de lógica borrosa", Proyecto Fin de Carrera, Universidad de Vigo, Vigo, 1997.

[NEURALOGIX 91] Fuzzy Microcontroller Development System Manual, American Neuralogix Inc., 1991.

[SUGENO 85] Sugeno, M., Nishida, M., Fuzzy control of model car, *Fuzzy Sets and Syste ms 16*, pp. 103-113, 1985.

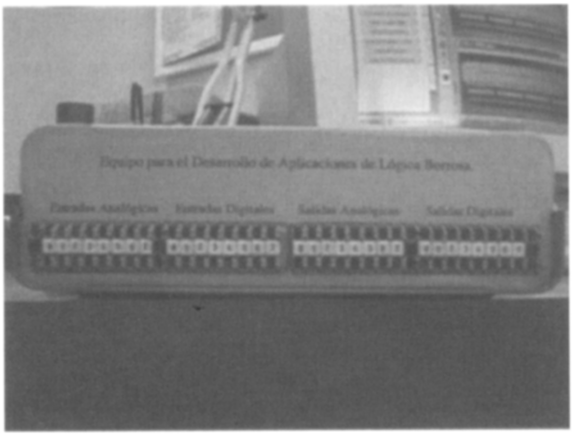

Equipo para el Desarrollo de Aplicaciones de Lógica Borrosa
Entradas Analógicas
Entradas Digitales
Salidas Analógicas
Salidas Digitales

Figure 7. Front of the equipment.

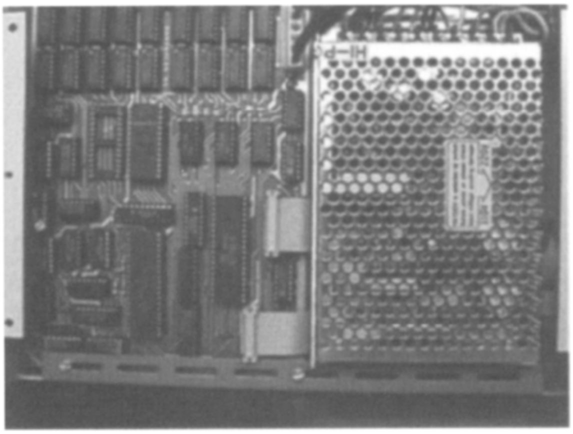


Figure 8. Inside view.

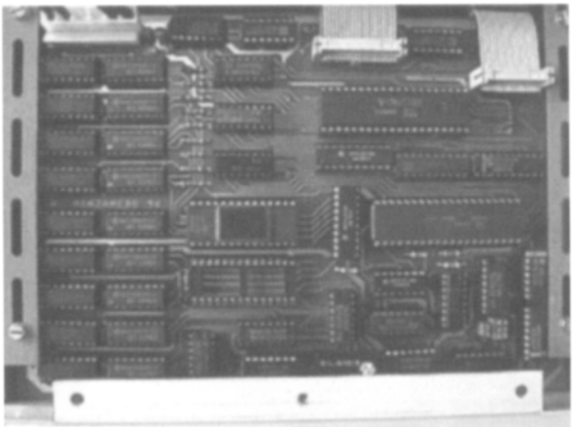


Figure 9. Detailed view.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Tribological Measurement of the State of Surface Fabrics by a Contact and a Non-contact Methods

Marie-Ange BUENO, Marc RENNER and Bernard DURAND

Laboratoire de Physique et Mécanique Textiles,
Ecole Nationale Supérieure des Industries Textiles de Mulhouse,
Université de Mulhouse
11, rue Alfred Werner, 68093 Mulhouse, France.

For the textile industry the touch of a fabric has become a very important selling argument ; this is why several processes in textile finishing meant to improve fabric touch are performed : sanding and raising, for example. It is now necessary to characterise these processes objectively, at present the control is very empirical. This study describes two methods for investigating sanding and raising and brings into light the effects of these processes on the fabric surface. Two different multi-directional roughness meters have been developed : a mechanical and an optical one. A textile fabric is rubbed with a probe or lighted with a beam and the signal obtained with each method is studied in the frequency domain. The power spectrum density, resulting to Fourier transform, shows several peaks of frequency which correspond to the kind of weave or knit and of the fabric density. The peak height varies and decreases after sanding and raising, due to a modified surface profile. The mechanical or optical multi-directional roughness meters provide information about the state of surface fabric and the basic directions of fabric relief.

1. SANDING AND RAISING

Sanding and raising are the most famous processes for the modification of the touch of fabrics. These are mechanical processes which modify the state of surface fabrics.

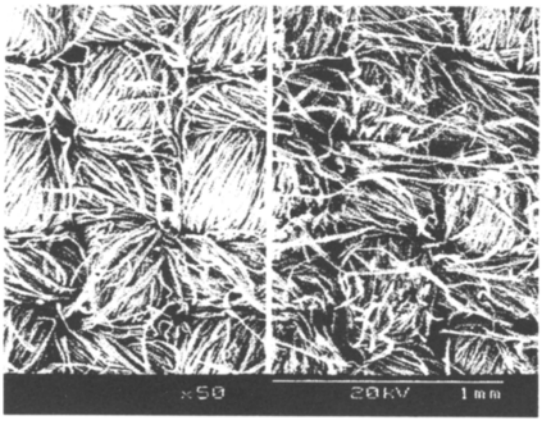


Figure 1. Fabric before (left) and after (right) sanding.

1.1. Sanding

Sanding is an abrasive wear that mainly modifies the touch and appearance of woven fabrics (peach skin touch, for example) [1,2]. In most cases, these fabrics are made from spun yarns.

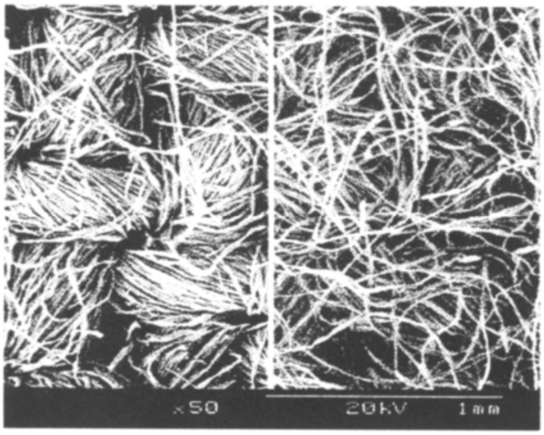


Figure 2. Fabric before (left) and after (right) raising.

Sanding is a surface wear obtained with rolls covered with emery slats or cylinders covered with an abrasive material that can be sand paper or diamond powder. It changes the fabric surface state, so that the look and the touch of the fabric will be modified. This process represents abrasive wear on two scales -the "microscopical" and "macroscopical" one.

Microscopical wear affects the fibres directly. After sanding, a fibre is strongly abrased towards its ends and on its surface. This abrasion can lead to fibre breakage, hence the presence of hairs on the textile surface.

Macroscopical wear affects the yarns (Figure 1). The surface hairiness of a fabric is due to fibres which are not contained within the body of a yarn. These fibres are linked to the yarn either by one of their ends or by both ends (in this latter case, the fibre forms a buckle). During the sanding process, the longest fibres are worn and/or broken. In a second step, the abrasive material wears the yarns (the weft or warp yarns, depending on the kind of weave), that is, it breaks the peripheral fibres without extracting them. For one worn fibre, two hairs are formed. The length of hairs depends on the kind of weave or knit, the fabric density and the sanding intensity. The length of hairs lies between 0 and the float length (the float being the piece between two interlacings). Therefore, the length dispersion is smaller and the number of hairs higher after sanding. The pile obtained is sequenced because of the small length dispersion and spatial disposition. Sanding gives short, but dense, regular hairs. In a woven fabric, for example, the warp is, in most cases, tighter than the weft. Therefore, in a plain woven fabric the tops of the asperities are formed by weft threads, and only these threads are worn. In the case of a warp twill woven fabric, the asperities are formed with warp threads, so that only these threads will be sanded.

1.2. Raising

Raising is a mechanical finishing process for woven and knitted fabrics, which can be performed in dry or wet conditions. It modifies the touch and appearance of fabrics, but the first objective is to increase the fabric's heat insulation ability. Indeed, raising creates a fibre mat in which air is confined, insulating the fabric from the environment.

Raising consists in brushing a fabric with wire-covered rollers or cylinders covered with teazles. Wet raising gives a flat pile and dry raising gives an erect pile. Raising affects peripheral fibres as the needles pull out some fibres from the yarn. So, the fibre mat formed on the fabric surface is quite thick and hides the fabric structure (Figure 2). The hair length distribution is greater than it is after sanding. Basically, the length of hairs is irregular because the hairs are pulled out to various extents. In most cases, the hairs are longer after raising than after sanding for the same fabric.

Raising therefore produces irregular hairiness in length dispersion and spatial disposition.

2. EXPERIMENTAL

2.1. A contact multi-directional roughness meter

The principle consists in rubbing the fabric with a probe (Figure. 3) and the signal obtained is the friction force or the vertical displacement according to the sensor used. The apparatus is a roughness meter including a sample-carrier, a sensor and the signal processing, representing a roughness meter (Figure 4). The sample-carrier has a rotating movement, then the fabric in all the surface directions can be measured [3].

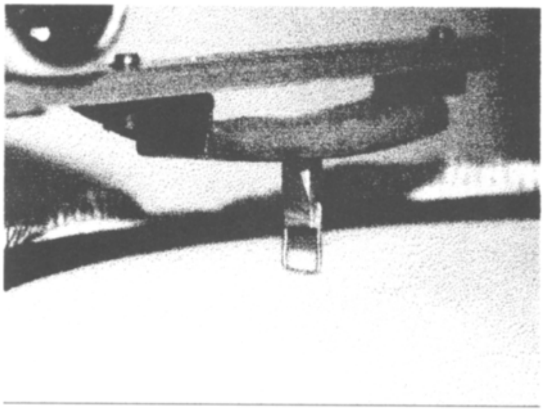


Figure 3. Probe of the mechanical roughness-meter

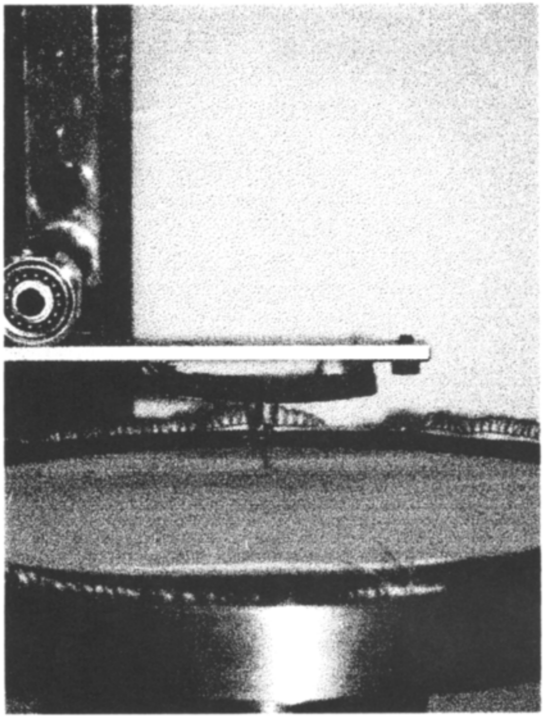


Figure 4. The mechanical roughness meter

To extract information from the temporal signal, rescued from the sensor, the Fourier transform and, more precisely, the power spectral density (or PSD) is computed :

$$PSD(f) = |X(f)|^2 \qquad (1)$$

where f: frequency in Hertz (s^{-1}).
$X(f)$: the Fourier transform of the temporal signal $x(t)$, which corresponds here to vertical displacement of the probe.
PSD : power spectral density.

The power spectral density is then the square modulus of the Fourier Transform of the signal versus the frequency ; the curve obtained is named autospectrum.

According to Parseval's theorem, the total energy E_X of a signal is not linked to the display in the time domain or in the frequency domain :

$$E_X = \int_{-\infty}^{+\infty} |X(f)|^2 \, df = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt \qquad (2)$$

2.2. A non-contact multi-directional roughness meter

The basic target is to direct a luminous beam on the fabric surface and to rescue the reflected beam. This beam included the information of the surface state.

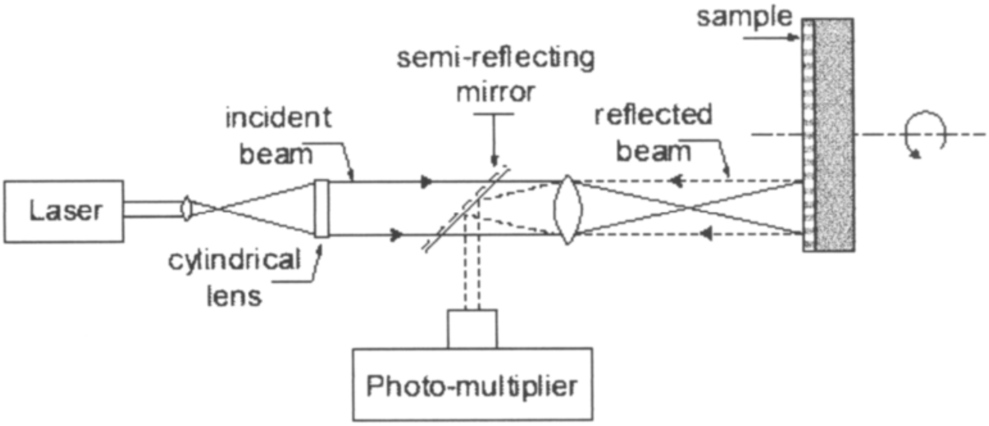
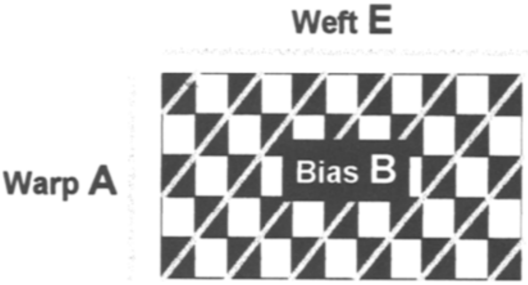


Figure 5. The optical multi-directional roughness meter

The optical multi-directional roughness meter includes three parts : the drive of the sample, the optical device and the signal processing. The sample-carrier has a circular movement, i.e. a multi-directional movement similar to that of the mechanical roughness meter. The optical assembly included a laser, because it is necessary to have a parallel beam to be accurate and some lenses (Figure 5). The light beam goes on the fabric surface and is reflected. The reflected ray is directed into a photo-multiplier which turns light into an electrical signal. The signal processing is the same as the contact method : the power spectral density (PSD) of the signal is computed by a spectrum analyser.

3. RESULTS

3.1. Sanding



As for example, a plain woven fabric is tested with both multi-directional roughness meters before and after sanding.

A plain woven fabric is a three basic directional structure (Figure 6), the bias, weft and warp directions. These directions are visible in the autospectra of the plain woven before and after sanding (Figures 7 and 8) because of the three peaks for both optical and mechanical multi-directional roughness meters.

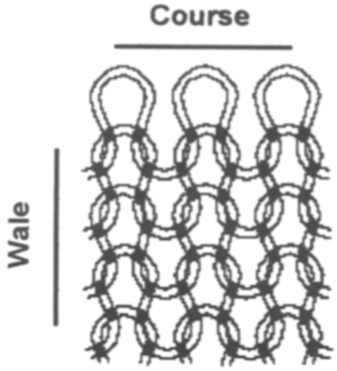


Figure 9. Knitted fabric.

After sanding, all the peaks of the autospectra decrease. Therefore, these measurements prove that both multi-directional roughness meters can make the difference between a fabric before and after sanding because of the lower peaks after process.

Even if the evolution of the autospectra is the

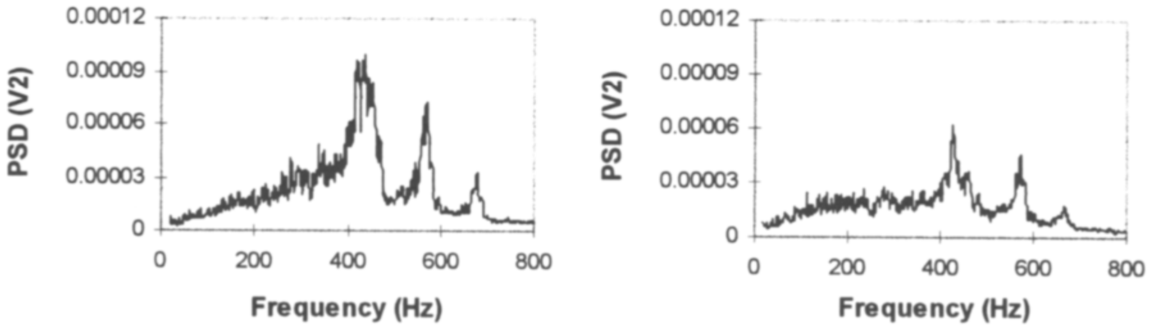


Figure 7. Mechanical roughness meter : autospectrum of the plain woven fabric - before sanding (left), after sanding (right).

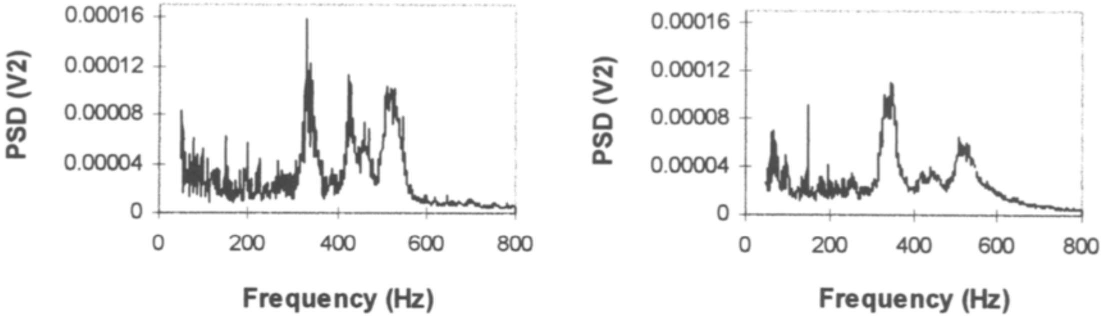


Figure 8. Optical roughness meter : autospectrum of the plain woven fabric - before sanding (left), after sanding (right).

same for both methods, they don't measure exactly the same thing. For the mechanical method, the wear of the fabric structure is characterised. In actual fact, the fabric is smoother after sanding. For the optical method, the pile due to the process absorbs a part of the reflected ray, then the total energy decreases and this is why the peaks of frequency are lower.

3.2. Raising

A knitted fabric is tested (Figure 9).

For this knitted fabric the column direction is the most important and in the autospectra before and after raising (Figures 10 and 11) the wale peak can be seen but after process it is lower.

In actual fact, the important hairiness due to raising hides the structure. In a first hand, the probe of the mechanical roughness meter is in contact with the fibre mat and not with the structure, therefore the structure peaks of frequency decrease. In a second hand, the energy absorption of the reflected beam is important as showing by the autospectrum after raising.

Both roughness meters are sensitive enough to distinguish a fabric before and after raising.

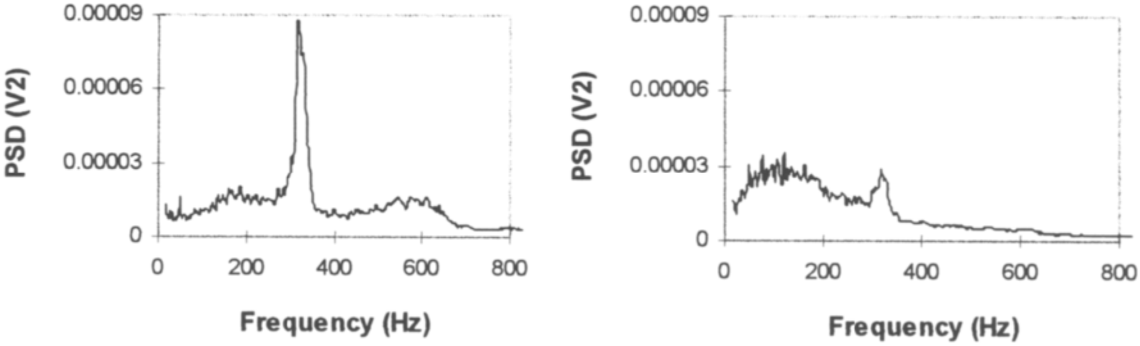
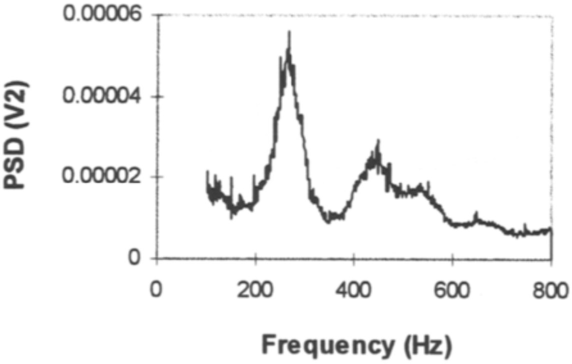
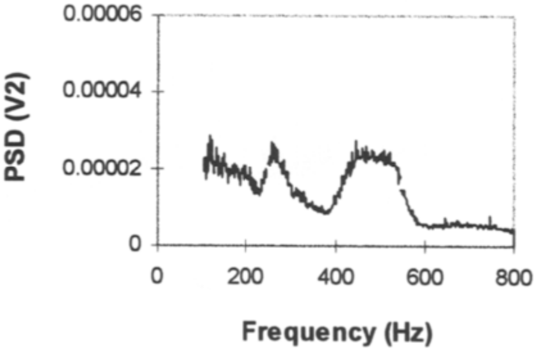


Figure 10. Mechanical roughness meter : autospectrum of the knitted fabric - before raising (left), after raising (right).





reflected, or a probe rubs the fabric and the profile of the surface is measured.

For the two methods presented in this paper, the signal processing is the same : the power spectral density (PSD) of the electrical signal obtained is computed. The curve representing the PSD versus frequency, named autospectrum, includes frequency peaks which represent the kind of weave or knit and the fabric density. The height of the autospectrum peaks depends on the state of surface fabrics.

Both processes, i.e. sanding and raising, are responsible for important hairiness on the fabric surface. Then after sanding and raising the peak height is lower. Therefore the mechanical and the optical multi-directional roughness meters have the ability to characterise sanding and raising processes, and others alike.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A novel mechatronic approach in machine vision for lace inspection

H. R. Yazdi[a] and T.G. King[b]
[a] School of Electronics and Electrical Engineering
University of Birmingham
Birmingham, B15 2TT, UK
h.r.yazdi@bham.ac.uk

[b]School of Textile Industries
University of Leeds
Leeds. LS2 9JT, UK
t.king@leeds.ac.uk

Flexible materials in general, and lace in particular, are hard to inspect using machine vision. Lace is especially difficult since it comprises fine and complex yarns that must be verified. For this purpose an inspection test rig using a mechatronic approach has been designed in order to simulate the production line. The main objective of this rig is to generate a *'comparable'* image. This has been achieved using an active vision system which uses the acquired image as *'feedback'* and hence creates a *'vision in the loop'* system.

1. INTRODUCTION

Decorative lace is a high value fabric which is mass produced in wide rolls. Lace is particularly difficult to inspect, using machine vision, since it comprises fine and complex yarns which must be verified. A major problem is that small distortions in the pattern are characteristic of the product and unavoidable. This is mainly due to the flexibility of the lace. Figure 1 shows a defect free piece of lace and figure 2 shows some of the faults that may occur during the manufacturing processes including those due to malfunction of the scalloping system (edge trimming and separating).

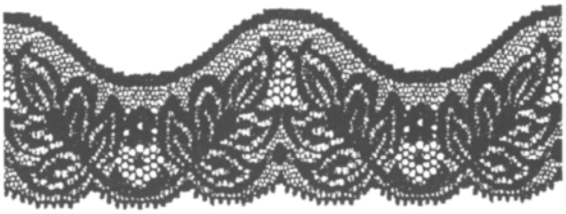


Figure 1. Defect free lace (reference image)

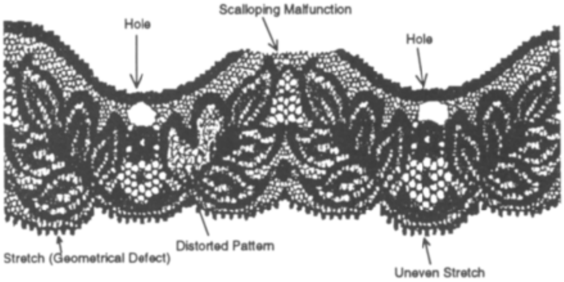


Figure 2. Defective lace, showing some of the possible defects

2. LACE INSPECTION

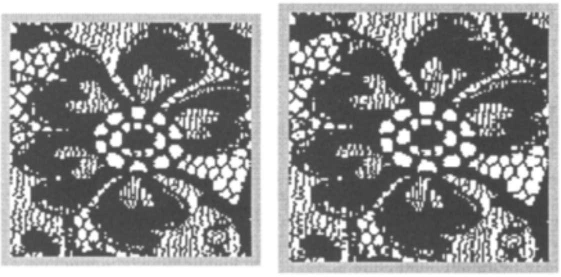
Previous research has indicated that some form of direct comparison of the lace being produced and a reference image is unavoidable [1, 2]. Other reported methods such as the use of the two dimensional Fourier power spectrum [1] are too computationally intensive, and therefore too slow, to be used in real time inspection implementations. King and Tao [3-5], have considered automating the process of lace scalloping. Sanby and Norton-Wayne [6] have worked on automated inspection of lace using machine vision. They realised that for detecting all types of possible detects (and not only hole-like ones), comparison between the pattern knitted and a perfect prototype is unavoidable. The major problem that must be taken into consideration is that due to the flexibility of lace, after subtraction, there will always be some kind of 'ghost' or 'residue' image.

For this reason Sanby and Norton-Wayne [6] exploited a trained Neural Net. The result of their approach was encouraging and led to less false alarms. It must be noted that this method can be used if, and only if, the lace is under tension and there are very small variations in the lace pattern. Presently, their method is being used while the lace is being knitted at very low speed (at around 10mm per minute). An attempt to utilise this approach after the scalloping process (i.e. after stentering and dyeing) was reported unsuccessful. [7]

3. STRATEGY AND TACTICS

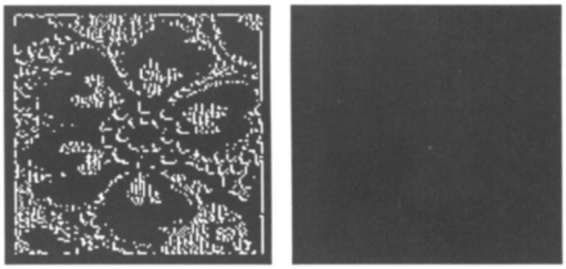
Due to the objectives of this research, it was decided to employ a direct comparison method. As mentioned earlier, we believe that due to the wide variety of lace patterns, this is the only practicable method. The problem of geometric variation in the lace being produced, either in size or orientation is a major barrier for direct comparison of two images. Figure 3 depicts this problem. It should be noted that the 'ghost' image consists of some 'residue pixels' (Salt Noise, noise that appears as a light spot on a dark background [8]). Later on in this paper, we have addressed this problem in more detail.

One possibility for solving this problem, could be 'software' manipulation of the image of the lace being inspected. For example, theoretically speaking, it is possible to 'shrink' the image in the example above and then directly compare it with the reference image (direct subtraction). But practically this process is very computationally demanding. Another feasible method is implementation of a mechatronic approach. The system is based on an 'Active Vision' concept which is not only 'flexible' but also 'intelligent' enough to 'align' itself with slight variations of the image that may occur during the image acquisition. Subtraction of the image of the lace being inspected and the model (resulting in the minimum amount of residue) can only be achieved when the two images have the same size and orientation. In this case, both 'random defects' such as holes, missing threads, etc. and 'geometric defects' such as stretch, skew, etc. will be detected. Hence the strategy is to 'make' the two images comparable and equi-size (isomorphing).



a) Original image
131x127 pixels

b) Stretch 10%
144x127 pixels



c) 'Ghost image'

d) Filtered image

Figure 3 (a)Original image of the lace; (b)The same image after stretching, using a commercial image processing software package; (c) 'Ghost' or 'residue' image: the result of direct subtraction of (a) and (b); (d) Morphologically filtered image.

4. THE EXPERIMENTAL RIG

A mechatronic approach has been designed. The test rig uses three stepper motors. The first motor controls the orientation of the camera, thus producing an 'aligned' image with respect to the model or reference image. The second motor controls the zoom factor of the lens, which adjusts the acquired image for possible stretch. The third motor controls the movement of the lace web, by rotating a transparent drum. Changing the relative rate of line scan image acquisition to drum movement allows longitudinal stretch to be compensated. The schematic of this rig and the concept of 'Vision in loop' is shown in figures 4 and 5 respectively.

All the required pulses for the stepper motors are produced using a specially designed I/O card connected to a PC. The software has been developed using the C++ language. The use of back lighting generates a clear image, which is converted to a binary image, in order to reduce the amount of data to be processed. In order to reduce the computational millstone, this process (thresholding or binarisation of the grey-level image) is achieved using a specially designed interface. The level of threshold can be selected dynamically by the operator. This card also is a means of synchronisation and buffering the data between PC and camera. The camera used is a 2048 pixel CCD line scan camera with clock frequency of up to 20 MHz (DALSA model CL-CX-2048G) [9].

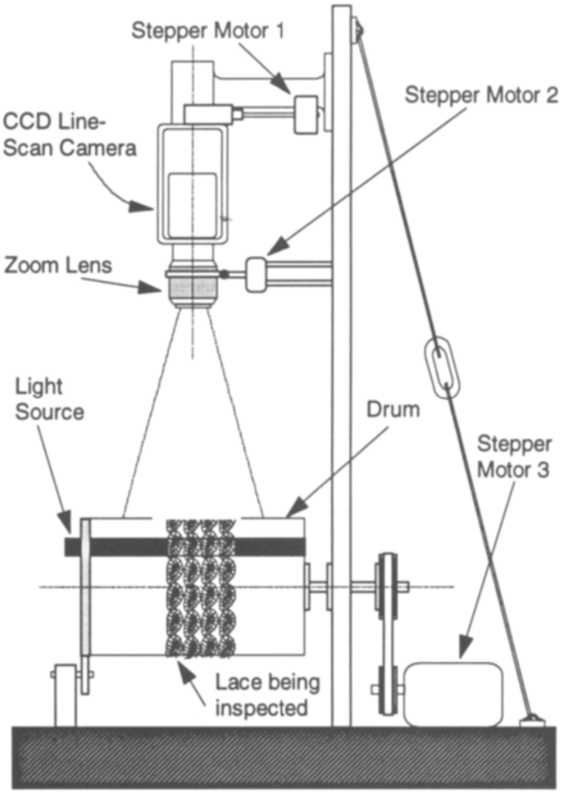

Stepper Motor 1
CCD Line-Scan Camera
Stepper Motor 2
Zoom Lens
Light Source
Drum
Stepper Motor 3
Lace being inspected

Figure-4 The designed test rig

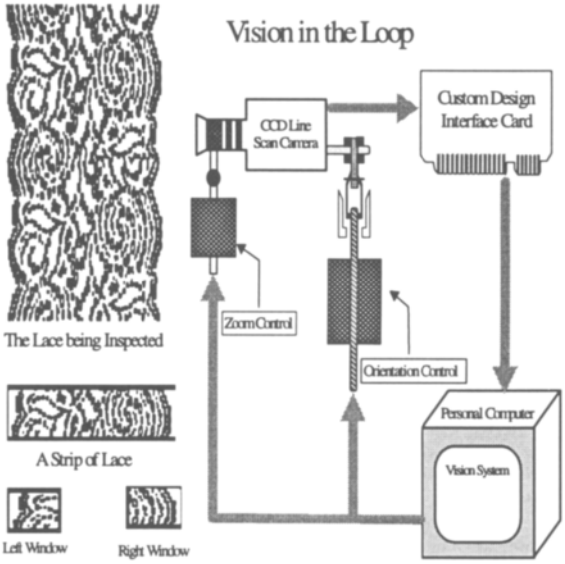

Vision in the Loop
The Lace being Inspected
A Strip of Lace
Left Window
Right Window
CCD Line Scan Camera
Custom Design Interface Card
Zoom Control
Orientation Control
Personal Computer
Vision System

Figure 5- Utilisation of camera as means for providing required feedback to control itself.

5. Vision In Loop, two 'windows' approach

The main objective is to compare the acquired image with a previously saved reference image. This can be achieved if the acquired and reference are 'comparable' against each other (both in size and orientation). If this is not the case, they must first be made 'comparable' and then perform the comparison. (i.e. subtract the acquired image or a part of it from the reference image) The alignment procedure consists of first, orientation alignment, which aligns the acquired image with the reference image in terms of 'skew' or 'rotation' and second, geometrical alignment which aligns the dimensions of the acquired image with the reference image. For this purpose, it was decided to acquire a strip of the image and select two portions (windows) of this strip, one on the left-hand-side and the other one on the right-hand-side of the strip. Then using either a correlation-based method or Distance transform-based method [10-12] find the best match within the reference image. This approach is illustrated in figure 6.

As can be seen in figure 6, after finding the best matches within the reference image, the required feedback for controlling the zoom and skew motors are easily calculated (δX and δY). The angle of skew has been calculated using:

$$\theta = \tan^{-1}(\delta x/\delta y) \equiv (\delta x/\delta y) \qquad (1)$$

Notice that in (1) the approximation is used for $-6° < \theta < +6°$

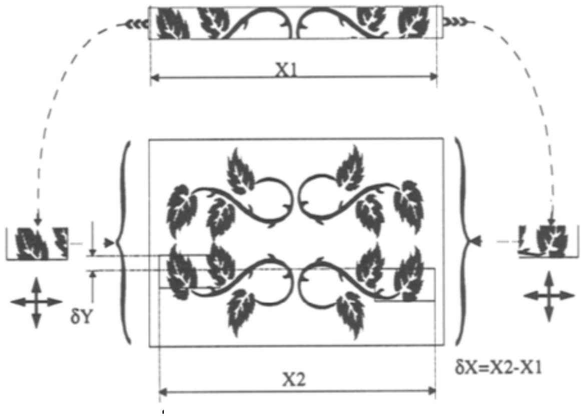


Figure 6- The concept of the ‘two windows’ approach, δX and δY will provide the required feedback

6. BINARY CORRELATION

A common method for finding the best match of a sub-image $t(i,j)$ of size WxH within an image $g(i,j)$ of size K x L where we assume W<K and H<L, is cross correlation function [15],

$$R(m,n) = \sum_i \sum_j g(i,j).t(i-m, j-n) \qquad (2)$$

where m=0,1,2,...,K-1, n=0,1,2,...,L-1, and the summation is taken over the image region where t and g overlap (region of Interest, ROI). For any value of (m,n) inside $g(i,j)$, application of (2) yields one value of R. As m and n are varied, $t(i,j)$ moves around the image area, giving the function $R(m,n)$. The maximum value of $R(m,n)$ indicates the position where $t(i,j)$ best matches $g(i,j)$. In a binary image, this will be an XORing process. For finding the best match, there are two alternatives, one is based on finding the m and n that **maximise** the following function (maximising similarities):

$$R(m,n) = \sum_i \sum_j \overline{g(i,j) \oplus t(i-m, j-n)} \qquad (3)$$

In practice (3) is very computationally demanding. For example, if the size of reference image is 300x500 pixels and size of the left window is 128x36, then the number of required XORs is: $128 \times 36 \times (300-128) \times (500-36) = 3.7 \times 10^8$.

An alternative is based on finding the m and n that **minimise** the following function (minimising the differences):

$$R(m,n) = \sum_i \sum_j g(i,j) \oplus t(i-m, j-n) \qquad (4)$$

In theory, (3) and (4) both provide the same result, but using (4) is much faster because we are searching for the minimum value of R (for each m and n and for all i’s and j’s); therefore at each stage, if the value of R exceeds its value at a previous location, then the XORing operation will be aborted (‘Early Termination’). This increases the speed of operation considerably. Hence in this work (4) has been used.

7. MORPHOLOGICAL FILTERS

After successfully finding the left and right windows, geometrical defects such as skew and stretch are easily detected and if they are not within the acceptable tolerance then the appropriate action needs to be taken. Only direct comparison of two images can reveal ‘Random’ flaws. In this case, the problem of ‘ghost’ or ‘residue’ image can pose a serious barrier, which might lead to numerous false alarms. Therefore the authors decided to somehow distinguish the residues from real defects. For this reason one must eliminate the ‘ghost’ image while preserving the defects. For this purpose Morphological filters i.e. Erosion and Dilation were used. [8, 13, 14]

8. QUANTITATIVE RESULTS

In our rig, the lateral and the longitudinal resolution are around 3.6 pixels/mm and 2.4 pixels/mm respectively. It should be noted that the longitudinal resolution is dictated by the mechanical structure of the test rig. Figure 7 illustrates around 14 cm of inspected lace. Figure 8 shows the result of direct comparison after morphological filters (Erosion and Dilation). Notice that 4 obvious defects were detected successfully and the system disregarded small discrepancies. The adjustable ‘sensitivity’ of the system governs the minimum size of detectable faults. In this example the minimum size of detectable defects is about 1mm x 1mm. The higher the sensitivity of the system the less the size of detectable defects but the higher the possibility of false alarms. In this example the system did not raise any false alarms

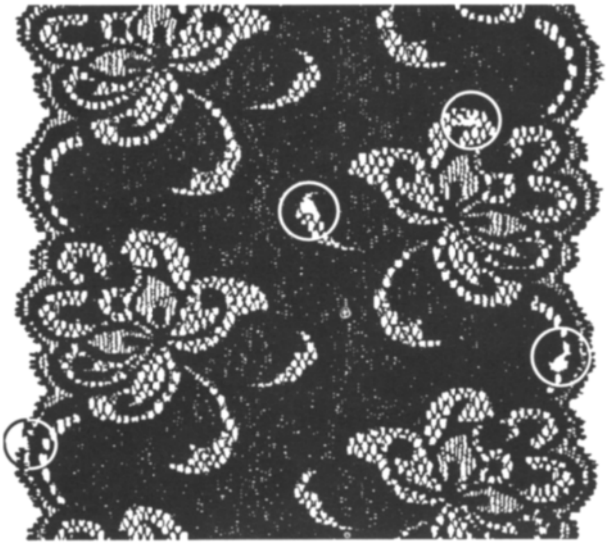


Figure 7. The defective lace and detected faults. (Shown in 4 circles)

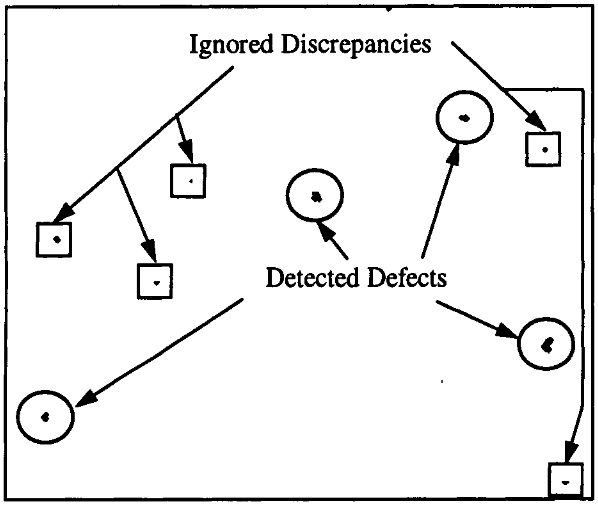


Figure 8 Detected defects and disregarded anomalies

8.1 inspection speed

Presently the speed of inspection of a typical sample lace is around 5.71mm/sec (0.342 m/min). The following pie chart shows the typical required time for inspection of 1 meter lace. It can be seen that calculation of 2 dimensional correlation is the most time consuming process. Hence, utilisation of dedicated hardware and parallel processing will enhance speed of defect detection considerably.

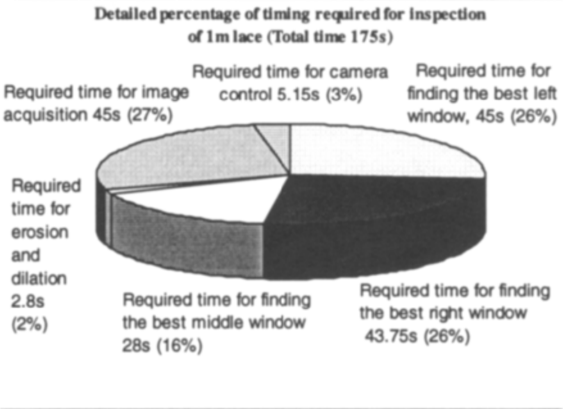


Figure 9. Detailed timing of the system

9. CONCLUSIONS

Automatic inspection of lace requires direct comparison of the image of the lace being inspected with an already acquired image of a reference lace. Comparison of two image is possible when two image have the same size and orientation (aligned). Using a flexible image acquisition system can overcome misalignments and will greatly enhance the result of comparison. The mechatronic approach is capable of coping with unpredictable distortions and 'adapting' itself to such small deformations. Using this system, we were able to achieve lace inspection successfully.

Morphological filters showed that they are capable of performing various image processing tasks and are a powerful tool for 'cleaning-up' processes. The speed of system is still under investigation, and we aim to use dedicated hardware and multiprocessing techniques in order to enhance the speed and accuracy. The utilisation of Digital Signal Processors and dedicated hardware Correlators is also being considered.

10. ACKNOWLEDGEMENTS

Support for this work from the UK ESPRC under grant GR/K 42561 'High Speed Processing of Deformable Webs: Computer Visual Inspection and Control', is gratefully acknowledged. Thanks are also due to Guy Birkin Ltd., Nottingham UK and Alan Shelton Ltd./Shelton Vision Systems Ltd., Leicester UK, for their support.

REFERENCES

1. Norton-Wayne, L., 'Inspection of Lace Using Machine Vision', Computer Graphics Forum 10, pp 113-119, 1991.
2. Sanby, C and Norton-Wayne, N., 'Machine Vision Inspection of Lace Using a Neural Network', Proc. SPIE, Vol. 2423, pp 314-322, 1995.
3. King, T.G. and Tao, L.G., 'An incremental real-time pattern tracking algorithm for line scan camera application', Mechatronics 4, No. 5, pp 503-516, 1994.
4. King, T.G., Tao, L.G., Jackson, M.R., Preston, M.E. and Yang, S., 'Computer vision controlled high speed laser cutting of lace', Proc. 2nd Int. Conf. on Computer Integrated Manufacturing (ICCIM '93), Vol.2, pp 929-936, Sept 6-10, Singapore, 1993.
5. King, T.G., Tao, L.G., Jackson, M.R. and Preston M.E., 'Real time tracking of pattern on deformable materials using DSP', Proc. IEE SERTA '93, pp 178-183, Cirencester, 1993.
6. Sanby, C., Norton-Wayne, L. and Harwood, R., 'The automated inspection of lace using machine vision', Mechatronics, Vol. 5, No. 2/3, pp. 215-231, 1995
7. Personal contact of the authors and visiting Shelton Vision Ltd. and Guy Birkin Ltd.
8. Davies, 'Machine Vision, theory, algorithms, practicalities', 2nd Edition, Academic Press Inc., 1997, pp 647-656.
9. Image Sensors and Cameras, Dalsa Inc. Waterloo, Ontario, Canada, 1997.
10. Yazdi, H.R. and King, T.G., 'Machine Vision for Lace Inspection Using Fuzzy Correlation and Distance Transformation', Proc. Int. Conf. on Intelligent and Cognitive Systems, September 1996, Tehran, Iran, ISBN 964-90010-3-4
11. Yazdi, H.R. and King, T.G., "Distance Transformations in Digital Images and their Application to Visual Inspection", Proc. Mechatronics '96 with M^{2}Vip '96, Vol. 2, pp 341-347, Guimarães, Portugal, 1996. ISBN 972-8063-08-3
12. Yazdi, H.R. and King, T.G., "Automatic Machine Vision for Lace Inspection", SPIE Photonics East '96: Machine Vision Applications, Architectures, and Systems Integration V, Proc. SPIE, Vol. 2908, pp 109-117, Boston, Mass. USA, 18-22 November 1996. ISBN 0-8194-2310-6
13. Heijmans, H.A.J.M., 'Mathematical Morphology: Basic Principles', Proceedings of Summer school of Morphological Image & Signal Processing, 1995, Paper available from Morphology Digest site via anonymous ftp to ftp.cwi.nl directory /pub/morphology/report/.
14. Heijmans, H.A.J.M., 'Mathematical Morphology: a Geometrical Approach in Image Processing', Nieuw Archief voor Wiskunde , Nov. 1992. Paper available from Morphology Digest site via anonymous ftp to ftp.cwi.nl directory /pub/morphology/report/.
15. Gonzalez, R.C. and Woods, R.E., 'Digital Image Processing', (2nd edition) Adison-Wesley, Reading, Mass., USA, pp 583-584, 1993.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Digital drive system for stitch length control in circular knitting

T. Dias, Tang Chien-Fa and G. Lanarolle

Department of Textiles, UMIST, Manchester M60 1QD

Abstract

A digital drive system has been created for driving the positive storage yarn feeders of multi-feeder circular knitting machines. This system is based on a brushless DC servomotor driving the feeders that supply the yarn into the knitting zone of a circular knitting machine. In a conventional positive feeding system it is still rather difficult to change the yarn feed length per needle (stitch quality setting) accurately. Currently in circular knitting machines the stitch length is altered by manually adjusting the diameter of the quality pulley. The digital drive system is controlled by a personal computer, and thus the yarn feed rate of the positive storage feeders can be changed easily, rapidly and precisely from the personal computer's keyboard.

The experimental results show that the system has a very high degree of accuracy and repeat-ability of yarn delivery rates.

1. INTRODUCTION

1.1. Background

On large diameter circular knitting machines, needles are placed in the tricks of a metal cylinder. Due to the rotational movement of this cylinder the needles are raised and lowered using non-linear cams. The size of the stitch and hence the dimensions of the knitted structure are influenced by the amplitude of the backward movement of the needles during stitch formation, and as such knitting machines are integrated with mechanisms that enable this amplitude to be adjusted. A machine with low maintenance and longer running times is a prerequisite for economical knitting. In circular knitting an important factor that influences the machine running time is the time needed for setting the machine to a different stitch quality.

This paper describes the work carried out at UMIST on the stitch length setting on a multi-feeder circular knitting machine by using a digital drive system to drive the positive storage feeders which is controlled from a standard personal computer (PC). User-friendly software has been created for setting the stitch quality of the machine easily, quickly and accurately.

The term 'digital drive system' refers to a microprocessor-based drive system, comprising a DC brushless servomotor, a shaft encoder and interface hardware. The system is flexible, easy to maintain and operate. The created software has the following user-friendly features:

- ❒ on-line control and adjustment of the yarn feed length per needle;
- ❒ the ability to down load important knitting parameters to the PC, which can be stored in production data files for future reference.

1.2. The current yarn feeding technology in circular knitting machines

1.2.1. Introduction

The production of high quality fabric is gaining importance because of increasing consumer consciousness and tough competition in the global market. On circular knitting machines positive yarn feeding devices play a major role in controlling the size and the regularity of the stitches formed. In fact the yarn-feeding device has become one of the most important elements of a modern multi-feeder circular knitting machine.

The main purpose a yarn-feeding device is to draw yarn from a package and deliver it to the needles to form stitches during a defined interval. Generally in circular knitting, this is applied at all yarn feeding points, i.e. at every knitting cam system of the machine.

1.2.2. Positive feeding devices for circular knitting machines

Isaac Rosen's tape nip yarn delivery is considered as the breakthrough in positive feeding for multi-feeder circular knitting machines. Here the yarn delivery to the needles was realised by passing the yarn between a moving tape and a rotating yarn feed wheel. The yarn feed wheel was driven by the moving tape. The yarn was trapped between the tape and the wheel, and it was driven to the needles at the same linear speed as the tape (Figure 1).

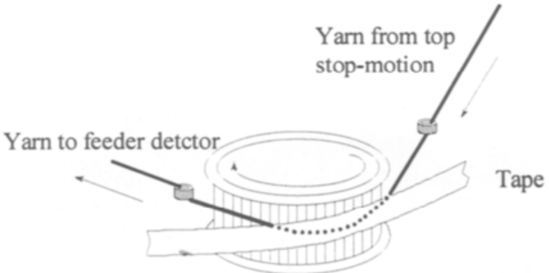


Figure 1. Tape nip positive yarn feeding

In order to equalise the feed rate of all yarn ends that were delivered to the needles at different cam systems, all the feed wheels of a circular knitting machine were driven by the same tape. The tape was surface-driven from a pulley with an adjustable diameter, known as the quality adjusting pulley, which was driven from the main drive of the machine (Figure 2). Adjusting the diameter of the quality pulley varied the tape speed, and the speed of the tape was changed according to the length of the yarn required to knit a course.

In order to reduce the weight of the yarn feed wheel and the accumulation of foreign particles on the yarn feed wheel surface, a squirrel cage type wheel was used. Here again due to the limited number of contact points between the yarn, tape and the yarn feed wheel pins, the needles were able to pull more yarn than delivered during knitting under higher run-in yarn tensions.

Another problem encountered during knitting was that where a yarn breakage occurred between the needles and the feed wheel, the broken yarn end was taken by the moving tape, and this prevented the yarn stop motion from being activated.

It was also found that delivering a measured length of yarn under higher production rates was difficult due to yarn slippage between the feed wheel (pinwheel) and the driver belt. This was due to the fact that the pinwheels were surface driven by the belt.

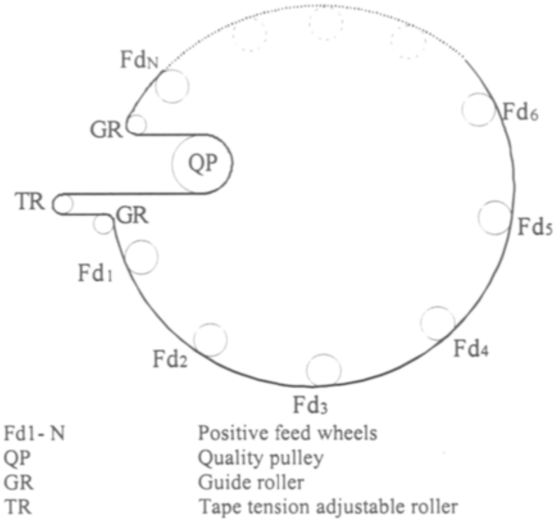


Figure 2. Schematic diagram of the positive feed System

In order to overcome the above difficulties a new type of positive feed system was developed in the 70's, namely the positive storage feed system. Here, a yarn feed wheel is driven by a separate lightweight plastic gear wheel/pin wheel fixed to the same shaft as the yarn feed wheel, and a tooth belt/punched belt is used to drive the gear wheel/pin wheel.

This arrangement eliminates slippage between the drive belt and the yarn feed wheel. It also allows the number of contact points between the feed wheel and the yarn to be increased by wrapping about 12-15 turns of yarn on to it. This creates a temporary yarn reserve immediately before the needles.

Memminger-IRO GmbH in Germany leads the world market in positive yarn storage feeder technology, and the most widely used positive feeder type of their product range, the MPF10-k1 is shown in Figure 3.

The advantages of storage positive feeder to circular knitting are given below:

(1) the magnetic brake ring provides self-cleaning and a maintenance-free vibration tensioner, which also has axis-free mountings.

(2) the recommended 12-15 wraps will give a yarn reserve of over 2m, which is a safety margin against empty yarn packages or yarn breakage before the yarn end reaches the knitting zone.

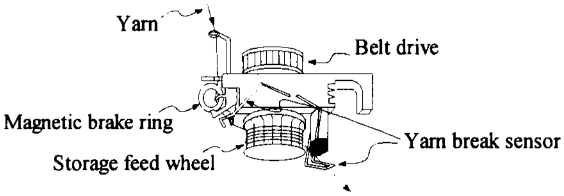


Figure 3. Positive storage yarn feeding device

1.3. Description of the Problem

In high production large diameter circular knitting machines, positive storage feeding has become the common standard. Although the concept enables the knitter to achieve a very high degree of uniformity in the product, it is still difficult and time consuming to adjust the machines to knit a specific stitch quality. This is mainly due to the fact that the yarn feed rate is determined by the tape speed which is mainly determined by the diameter of the quality pulley. As the feeder tape is surface driven by the quality pulley, the tape tension and the tape surface condition will also influence its speed. Changing the stitch length is achieved by adjusting the diameter of the quality pulley.

To achieve the required stitch length, it may be necessary to try several diameter variations, and waste time in adjustment. In addition, the results are usually influenced by an increase in stitch length variations caused by the belt's friction originating in the elongation of the belt and the gap in the segments of the quality pulley. Today with short changeover times, lead times and short-lived fashions, every step of knitted fabric production need to be rationalized.

2. DESCRIPTION OF THE UMIST DIGITALDRIVE SYSTEM FOR POSITIVE STORAGE FEEDING

2.1. Conventional yarn delivery system

A block diagram representation of a conventional yarn delivery system in multi-feeder circular knitting machines is shown in Figure 4.

Conventionally, changing the stitch length in circular knitted fabric involves altering the diameter of the adjustable quality pulley. The belt is surfaced driven by the quality pulley, and therefore slippage between the belt and the quality pulley must be prevented to minimize course length variations in the knitted fabric. The belt slippage depends on the surface condition of the quality pulley and the belt as well as on the belt tension. The belt tension alters with use due to belt extension and wear. Due to above reasons it is difficult and time consuming to readjust the quality pulley to a previous setting.

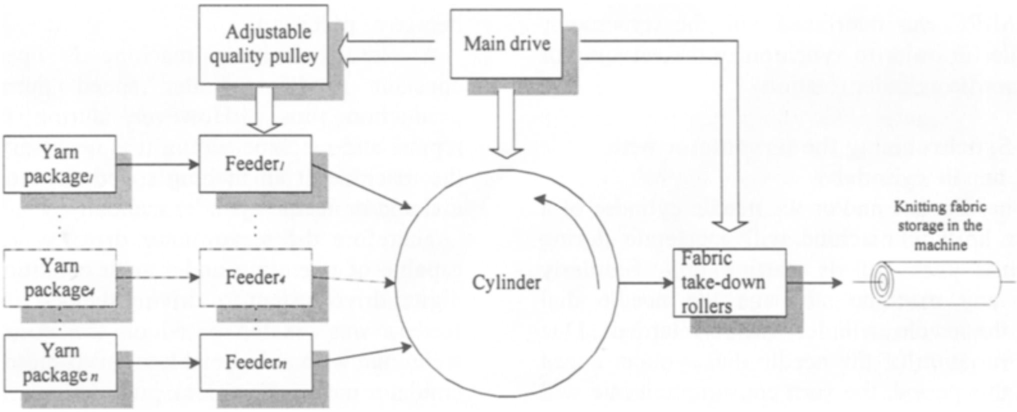


Figure 4. Conventional yarn feeding system in a circular knitting machine

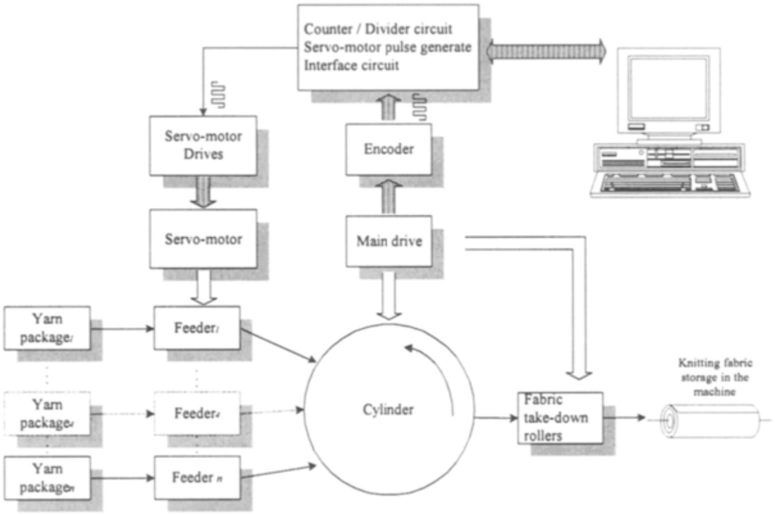


Figure 5. The new digital drive system on a circular knitting machine

2.2. The UMIST concept

A new concept was created for driving the belt that drives the positive storage feeders on a multiple-feeder circular knitting machine with a high performance dc brushless servo-motor, and its block diagram representation is given in figure 5.

A shaft encoder was attached to the needle cylinder drive mechanism of the knitting machine, and the pulses generated by it were used to synchronize the servomotor with the needle movement. In order to be flexible with the stitch length (course length) adjustment, it was decided to operate the servomotor in stepper emulator mode with suitable electronic control hardware. An IBM-PC was interfaced with the servomotor controller in order to synchronize the servomotor to the needle cylinder rotation.

2.3. Synchronizing the servomotor with needle cylinder

The needle dial and/or the needle cylinder of a circular knitting machine will accelerate during the initial phase of its starting up. Similarly during any machine stoppage the needle dial and/or the needle cylinder will be retarded. Due to the variation of the needle dial/cylinder speed during this period, the yarn consumption rate will change and the digital drive system must be capable of responding to such variations.

There are two popular methods for synchronizing rotational movements, i.e. the velocity control and the position control methods.

In the velocity control method a velocity transducer signal (e.g. a taco-generator) is used to generate a signal that is proportional to the speed of a shaft or a needle cylinder to be synchronized. This signal is compared with a reference signal, and if there is any difference then the speed of the motor is corrected with control/power circuits.

The position control method uses a different feedback signal to the velocity control method. The signal from a position transducer is used to generate electrical pulses to drive a stepper motor or a servomotor in step emulation mode with a higher degree of precision.

A circular knitting machine is operated at a constant needle cylinder speed during normal production runs. However, during faultfinding, repairs and machine setting it is necessary to operate the machine at an inching speed or turn the needle dial and/or needle cylinder manually.

Therefore the servo-motor drive system must be capable of operating under such conditions, and the digital drive system for driving the belt of the storage feeders was created based on the position control technique with a dc brushless servo-motor in stepper emulator mode. The clock pulses required to operate the servomotor were generated with specially designed hardware in conjunction with software to

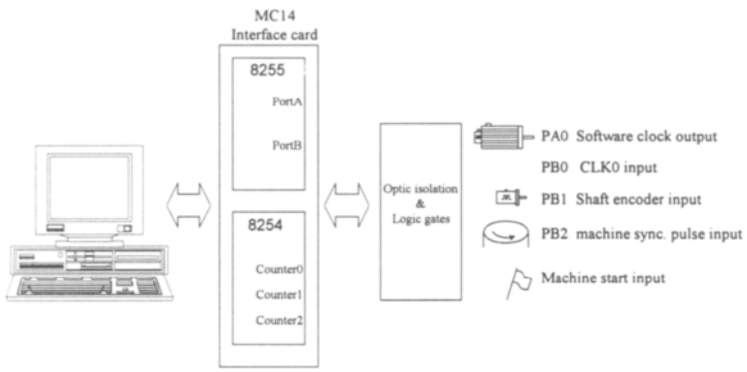

MC14
Interface card
8255
PortA
PortB
8254
Counter0
Counter1
Counter2
Optic isolation
&
Logic gates
PA0 Software clock output
PB0 CLK0 input
PB1 Shaft encoder input
PB2 machine sync. pulse input
Machine start input

Figure 6. The block diagram of interface circuit

improve the user-friendliness of the new design. The needle dial and/or the needle cylinder position, i.e. the needle position, was determined with the help of a shaft encoder installed on to the needle cylinder drive mechanism. The new concept is shown schematically in Figure 6.

Three counters were needed to synchronize the servomotor to the needle cylinder of the circular-knitting machine (Figure 7). When the circular knitting machine is running, the pulses from the shaft encoder are divided by the counter0 (N0). Counter0 outputs a signal to counter1, D type flip-flop and counter2. In order to synchronize to the 4MHz clock of the crystal oscillator, counter2 is enabled by counter0, and counter2 outputs a square wave which is a ratio of N2 to the 4 MHz clock. Counter1 starts counting the pulses from counter2 and output pulses after the counter0 trigger signal (CLK0) has transitioned from low to high.

Counters will generate pulses until the terminal count (N1) is reached. Prior to trigger by the CLK0, the D type flip-flop output will be low and force CLK to be low. After the CLK0 rising edge the D type output will go high, allowing CLK2 pulses to pass to CLK until the D type flip-flop is reset by counter output. CLK are the drive pulses to the servomotor drive unit. The servomotor will therefore synchronize with the shaft encoder signal while needle cylinder of the circular knitting machine is rotating. Figure 8 shows the control timing diagram.

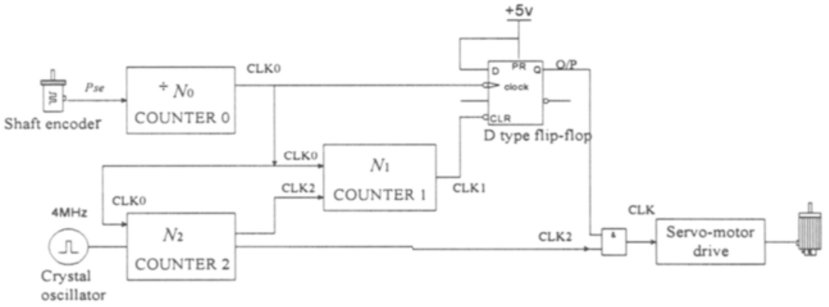

+5v
Shaft encoder
P_{se}
$\div N_0$
COUNTER 0
CLK0
D
PR
Q
O/P
clock
CLR
D type flip-flop
CLK0
N_1
CLK2
COUNTER 1
CLK1
4MHz
CLK0
N_2
COUNTER 2
Crystal
oscillator
CLK
CLK2
Servo-motor
drive

Figure 7. The block diagram for digital drive system and shaft encoder synchronisation

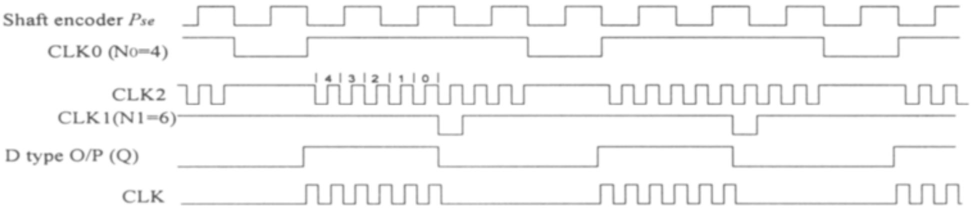

Shaft encoder P_{se}
CLK0 (N0=4)
4 3 2 1 0
CLK2
CLK1(N1=6)
D type O/P (Q)
CLK

Figure 8. The timing diagram of the control circuit

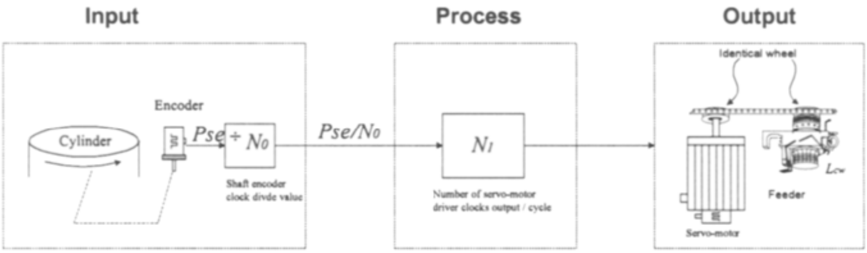


Figure 9. The block diagram of control model

2.4. Control model of digital drive system

In order to prevent slippage between the tape and its drive wheel the tape was driven by a wheel that is identical to the one being used to drive the yarn feed wheel of a positive storage feeder was used. The course length will depend on the resolution of the servomotor and the circumference of the storage feed wheel. The control model of the digital drive system is shown in figure 9.

The control model is given by the following mathematical relationship:

$$N_1 = \frac{SMres \cdot N_0}{L_{cw} \cdot Pse} \cdot Lsl \cdot Nn \qquad (1)$$

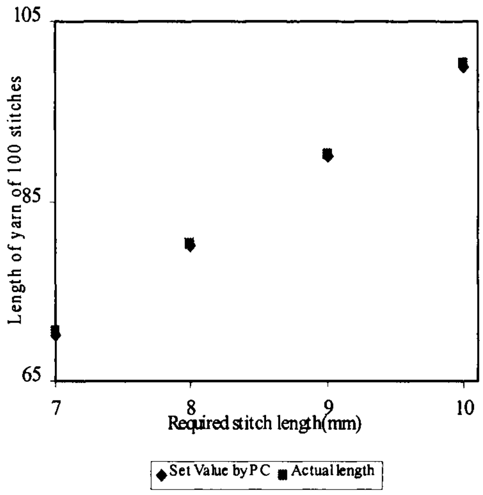
where

N1	The number of pulses needed by the motor controller unit to operate the servomotor to deliver a yarn length Lsl to the needles,
N0	The shaft encoder clocks divide value,
Nn	The number of needles in cylinder,
Lsl	The stitch length,
Lcw	The circumference of feeder's wheel,
Pse	The total number of shaft encoder pulses per cylinder revolution,
SMres	The resolution of the servomotor driver.

3. EVALUATION OF THE DIGITAL DRIVE SYSTEM

The new digital drive system was installed on a Camber DEPANIT, E7, 630 needles, and 24 feeder machine. In order to evaluate the performance of the new digital drive system a range of fabrics with different stitch lengths were knitted. The length of yarn of 100 stitches of fabrics knitted on the machine for a range of stitch lengths was measured in accordance with BS 5441. The results demonstrate the accuracy of the new system, see Figure 10.

Figure 10. Analysis of the effectiveness of the system



4. CONCLUSION

The digital drive system for controlling the positive yarn feed rate on circular knitting machines created at UMIST offers a greater flexibility for setting circular knitting machines to produce knitted fabrics of defined stitch quality. As the system is based on electronic pulse generation, it confirms the product reproducibility. An average deviation of about 0.4% from the required value is in fact a greater achievement when processing textile materials.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Mechatronic principles for aesthetic measurement of textile materials

G. K. Stylios

Centre for Objective Measurement and Innovation Technologies,
University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom

Textile materials are bought by consumers on the merit s of their aesthetic appeal. This can be measured and interpreted in terms of newly defined aesthetic attributes which can be engineered into the material during its design and production process. In the design, making up, selling and wear of apparel, two easthetic characteristics are very important; drape in standing (static) and in motion (dynamic), and deformation of the material along its sewing line (seam pucker). Both techniques are based on dedicated machine vision systems and are using lasers and/or CCD cameras in capturing aesthetic data.

1. GENERAL INTRODUCTION

Fierce competition, surplus of clothing and lack of own technology, render textile manufacture one of the most vulnerable industries. Short termism, survival rather than growth, sourcing against self sufficiency and wastage rather than maximisation of resources are synonymous with the textile chain. Despite this gloomy image of textiles however, there are models, of strategies which seek to effectively implement technological solutions to those problems with considerable success. Recent investigations which underpin new manufacturing innovative concepts are those emphasising textile aesthetic properties rather than properties related to the old principle of quality performance, and are trying to provide realistic means of measurement, interpretation and ultimately engineering of textile aesthetics as consumer attributes.

2. MECHATRONIC PRINCIPLES FOR AESTHETIC MEASUREMENT OF FABRIC DRAPE

Drape is considered to be one of the most important aesthetic attributes in textiles, because it is the main contributory factor in the designing and retailing of apparel products. The subject of fabric drape has been studied by a number of researchers[1], and the most widely accepted method of measuring fabric drape is by using the so called “drapemeter”[2]. This approach is based on a quantity known as the drape coefficient F which is a measure of the projected area of the fabric sample to its undraped area, supported by a disc, in which the area of the supporting disk is deduced. This method has some obvious shortcomings because it cannot provide an easthetic measure of the material, neither can determine accurately its drape behaviour. For example, both fabrics in figure 1 may cover the same area but have different aesthetic human assessment. This shows that drape coefficient cannot define fabric drape accurately.

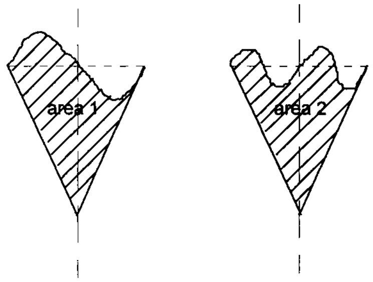


Figure 1. Area 1 = Area 2, Drape 1 ≠ Drape 2

Furthermore, fabric drapability under "dynamic" conditions, as in, the influence of air during twirling is also important, especially nowadays that garments are made with lighterweight fabrics. Consequently useful parameters for 3-D aesthetic characterisation in both static and dynamic conditions need to be made available.

Consequently after careful study of human perception in characterising drape [1], further parameters were introduced to quantify and describe the drape performance of textile materials. These parameters are based on cognitive psychology and are called aesthetic drape attributes; drape coefficient, the number of folds, the depth of folds and the eveness of folds.

2.1. Static and dynamic measurement of fabric drape

When human eyes see a garment, they gain a measure of its' drape from the relative differences between the bright and dark folds of the fabric. The use of a camera and a computer is realistic, in this case, since it simulates the human visual process of the eye and the brain in the assessment of garment drape. In the case of dynamic drape, it is generally accepted that the apparel article floats outwards (twirls) under centrifugal acceleration, reversal or other conditions imposed on it by the wearer or by other external forces such as air. To achieve effective measurement of aesthetic drape attributes, a dedicated system was devised called the M^3 Drape Measurement System. Figure 2 shows the different parts of the measurement system, as follows:

1. a CCD camera to capture the image
2. a cabinet with appropriate light conditions
3. a PC to extract the information from the image and to process the data
4. a monitor to display the image
5. a motorised supporting disk to hold the specimen and also to rotate the fabric for measuring the dynamic drape. It is possible to change the speed or incorporate reversal modes by changing angle directions of motion. The current mode was 43and 86 rpm.

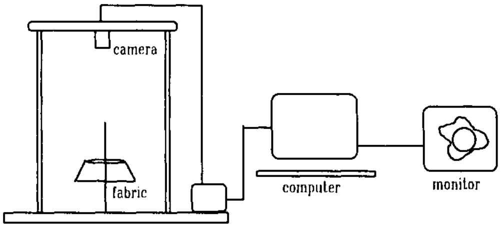


Figure 2. The M^3 Drape Measurement System

2.2. The Draping Algorithm [1]

In digital image processing we are able to define well, a variety of object classes. Object classes have qualities which are extracted from the captured images and referred to as P features. Each pixel or object is then represented as a feature vector in this space. If the features represented an object class well, all feature vectors of the objects from this class should lie close to each other in the feature space. Furthermore an object in feature space may be represented by its centre of mass. In this case the "minimum distance method" can be used to determine to which class a feature vector belongs to by computing the distance of the feature vector of all cluster centres and choose the class which has the minimum distance.

For pattern classification [3], feature vector **V** has to be introduced to recognise the different pattern, the number of parameters in the feature vector will depend on the details of the pattern which have to be considered. From a circular pattern with fixed outline length, three parameters will help us to classify the pattern, shown in Figure 3.

Feature vector V as: $= (\bar{\rho}_{\max}, \bar{\rho}_{\min}, S)$

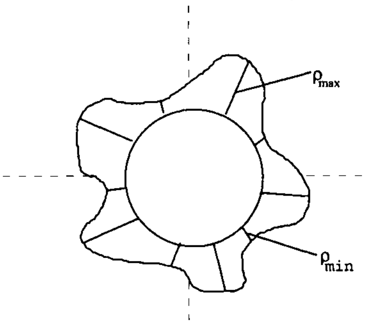


Figure 3. The projected area of a drape specimen

where: $\bar{\rho}_{\max} = \sum_{i=1}^{n} \frac{\rho^{\max}(i)}{n}$

$$\bar{\rho}_{\min} = \sum_{i=1}^{n} \frac{\rho^{\min}(i)}{n}$$

$$S = \sum_{i=1}^{n} \frac{(\rho_{\max}(i) \quad \bar{\rho}_{\max})^2}{\bar{\rho}_{\max}^{\,2}}$$

S represents how balanced or even the folds are and is calculate as a normalised variance of the maximum fold length from the average maximum fold length. Consequently, $S = 0$ when the folds are even and $S = 1$ if the variation in the fold length is of the order of a fold length itself.

A more intuitively useful quantity is angle α, Figure 4. This angle can be given using the trigonometric identity for sine;

$$\alpha = \sin^{-1}\left(\frac{\rho_{\max}}{i}\right) \qquad \text{n - number of nodes}$$

If we look at the 3-D fabric drape, the slope α is a much better representation for the fabric image, where : $\rho\max \Rightarrow \alpha\max$ $\quad \rho\min \Rightarrow \alpha\min$

So the feature vector may be as follows: $= (\bar{\alpha}_{\max}, \bar{\alpha}_{\min}, S)$

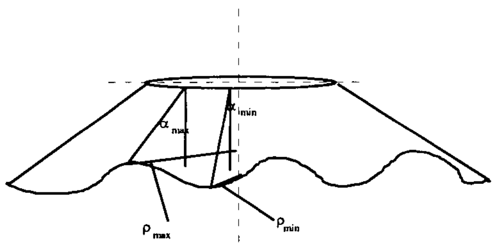


Figure 4. The Geometry of Fabric Drape

2.3. Experimental results and implementation

Five different fabrics have been measured in accordance with the draping coefficient and feature vector. The subjective judgement is as follows; Fabric 2 is draped highest, next is fabric 5, and the lowest is fabric 4. Table 1 shows the results of the objective measurement.

If only the drape coefficient is considered as shown in the above table, then fabric 2 has the highest drapability and fabric 3 has the lowest, which is different from the subjective assessment. In addition information about the actual deformation cannot be given. The fabric dynamic drape is also very important. It is possible that two fabrics have the same static drape parameters, but dynamically they can show significantly different drapability. For example, a heavy and stiff fabric may show the same static drape as a light and soft fabric, but dynamically these two show obviously different drape behaviour. Table 2 provides the dynamic drape measurement results of five different fabrics under two different rotating speeds. It can be seen that the difference between static and dynamic drape of fabric 2 is largest, so it possesses a very high dynamic effect. Fabric 3 is the most stable one, which means it is stiff and has low dynamic drape.

2.4. Discussion and conclusion

A "draping algorithm" has been devised from human perception to measure "aesthetic drape attributes" such as; drape coefficient, number of folds, depth of folds, and eveness of folds. Static as well as dynamic drape measurements have been successfully implemented on commercial fabrics.

Table 1 Objective measurement of fabric drape

fabrics	1	2	3	4	5
α_{max}	63.84	47.54	64.83	66.81	63.64
α_{min}	30.65	26.44	35.88	31.01	31.66
variation	0.0084	0.0495	0.0046	0.0036	0.0122
drape c.	0.3790	0.2233	0.4359	0.3653	0.3816

Table 2 The measurement of dynamic drape at speed 43rpm

		speed*=43			
fabrics	1	2	3	4	5
α_{max}	66.45	55.76	65.88	67.85	64.29
α_{min}	34.62	27.22	37.18	35.30	33.20
variation s	0.0134	0.0064	0.0049	0.0058	0.0108
drape c.	0.4209	0.2953	0.4551	0.4231	0.3958
		speed*=86			
α_{max}	78.56	73.99	75.44	81.79	73.45
α_{min}	46.54	41.14	51.67	51.01	56.62
variation s	0.0038	0.0048	0.0083	0.0031	0.0056
drape c.	0.5879	0.5468	0.6697	0.6615	0.6393

* the sample rotating speed(rpm). The speed can be selected easily according to different requirements.

2.4. Discussion and conclusion

A “draping algorithm” has been devised from human perception to measure “aesthetic drape attributes” such as; drape coefficient, number of folds, depth of folds, and eveness of folds. Static as well as dynamic drape measurements have been successfully implemented on commercial fabrics.

3. MECHATRONIC PRINCIPLES FOR AESTHETIC MEASUREMENT OF SEAM PUCKER

The deformation of materials during processing is one of the problems that the manufacturing industry has been facing for many years. The effect that severe seam pucker can have in a garment may be aesthetic or/and functional. The industry being aware of all these problems uses a set of photographic standards produced by AATCC[4] to assess whether the seams in garments are acceptable or not. This is an important measure of the aesthetic appeal of the quality of the seams of a garment to the customer, however problems arise when judges often disagree in the ranking of the sewn strips of fabric, and thus quality and specification standards are difficult to be implemented due to subjectivity and time involved in the ranking procedure.

There have been attempts in the past to mechanise the assessment of seam pucker, these techniques relied considerably on repetitive sampling, most of them involving contact between sensor and specimen for a single measurement, significantly undermining the accuracy and reproducibility achievable with the instruments [5], [6].

This paper investigates the role played by the cognitive process of a trained observer in seam pucker subjective assessment procedures. This role is then modelled by defining the parameters involved in the cognitive process using mathematical expressions, and implemented in a dedicated vision system described.

3.1. The seam pucker algorithm

With very few exceptions a sewn garment with nicely flat or curved seams is regarded as desirable, because the design line of the garment and its smooth "unpuckered" three - dimensional configuration, are the most predominant features of it. When the human eyes see a garment, they gain a measure of the severity of seam pucker largely from the relative difference between the bright and dark parts of the deformed (wrinkled) surface of the fabric along the seam line [7].

Borrowing terminologies from wave theory, "pucker wavelength" is defined as the distance between two adjacent troughs measured along the the plane of the seam, while "Pucker amplitude" is defined as the distance between a peak and an adjacent trough measured along a direction perpendicular to the plane of the seam. This methematical treatment has been reported by Stylios and Sotomi elsewhere [6].

It has been generally observed that if puckering is viewed as a waveform in the plane of a seam, puckering of increasing wave frequency is accompanied by decreasing seam quality while, puckering of decreasing wave frequency has an opposite effect. This visual assessment can be expressed in terms of "Pucker Severity" (P.S.) as,

$$\text{P.S.} = f(y),$$

where Y is the wave amplitude,

While the visual assessment of figure 2 can be expressed in terms of P.S. as,

$$\text{P.S.} = f(x),$$

where X is the wave length.

The combination of these two expressions gives,

$$\text{P.S.} = f(y/x)$$

If pucker severity is defined as the value of the ratio *y/x* per unit pucker wavelength or distance measured along the plane of the seam,

$$\text{P.S.} = \text{f}(y/x^2)$$

This is in agreement with observations from cognitive psychology which suggests that pucker severity decreases with increasing *x* for individual puckers with identical *y/x* ratio, since there is very little possibility that all the troughs in a seam would lie on the same plane, it is convenient and probably more accurate in practice, to determine P.S. for half puckers rather than full puckers. This means that "pucker amplitude" is unchanged but "pucker wavelength" is redefined as the distance between a peak and an adjacent trough, measured along the plane of the seam [6].

A "Pucker Grade Look Up Table" is constructed from the linear regression analytical results of obtained pucker indices and subjective assessments of trained judges, from which Pucker Indices are automatically assigned to AATCC photographic standards.

3.2. Aesthetic Measurement of seam pucker; Pucker Laser Measurement System PLMS

The PLMS$^{\text{TM}}$ figure 5, consists of a box-like cabinet inside which the laser arrays, the camera and the sample holder are situated. The box has an opening from where a sample slider (ruler-like) can be pulled out and inserted in for testing. The slider is also used as a fabric sample template for cutting fabric samples. The laser diodes (class 2) and CCD camera constitute the vision acquisition sensors of the PLMS. The computer program written in C++ enables operation of the system in which the image is reconstructed together with the severity deformation distribution of the measured sample,

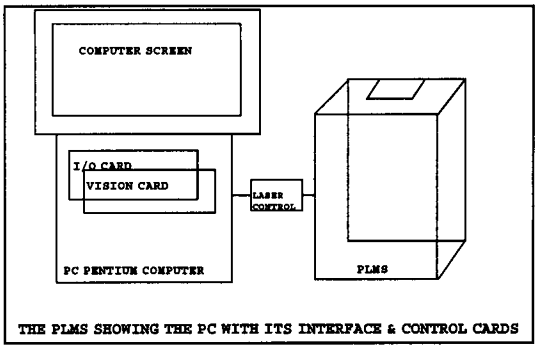


Figure 5 A schematic drawing of the PLMS

The box is connected with a vision interface card, with a laser control card (outside the PC) and an input/output card which are installed inside the PC and are occupying two of its expansion slots. The instrument is capable of measuring samples that are stitched in the middle or nearer to the edge of the fabric. Each complete measurement takes 2 minutes, for sample preparation to the completion of the test and the output is given in AATCC grades 1-5 or in-between for higher accuracy; 1.5, 2.5,3.5, 4.5.

3.3. Experimental Results

Twenty commercial lightweight mainly plain woven fabrics were assessed by the PLMS and by two judges. Only two fabrics were not in agreement with the PLMS but only by one grade, and there were no problems with dark and patterned fabrics, as shown in Table 3.

Table 3 Experimental results of seam pucker measurement

Fabric	Judge	PLMS	Fabric	Judge	PLMS
1	1	1	11	3	3
2	4	3	12	2	2
3	2	2	13	1	1
4	5	5	14	4	4
5	3	3	15	5	5
6	3	2	16	2	2
7	2	2	17	1	1
8	4	4	18	1	1
9	5	5	19	3	3
10	2	2	20	1	1

3.4. Discussion and conclusion

This paper reports upon findings of a new research approach for the measurement of seam pucker as an aesthetic property. The paper describes a cognitive model for the assessment of seam pucker which provides the foundations for the development of a new objective measurement tool based on computer vision. The PLMS is a dedicated measurement system devised for the accurate assessment of seam pucker and has overcome problems with dark, light and patterned fabrics.

4. GENERAL DISCUSSION

This paper has put forward some new concepts for the measurement of textile aesthetics which are important aspects in providing the textile manufacturing industry with technological alternatives to combat competition. This research which is part of "Intelligent Textile and Garment Manufacture" has already continued successfully in the engineering of textile attributes by using non-conventional models to link manufacturing parameters with aesthetic attributes , so that these properties can be reverse engineered [8] into new products Right First Time under demand led operations. It would have taken much more space to report on those findings here but in this paper the fundamentals of aesthetic measurement have been provided.

ACKNOWLEDGMENTS

I would like to acknowledge J O Sotomi, N Powell, R Zhu and T Wan who have been involved in this research effort.

REFERENCES

1. G.Stylios and R Zhu, J. Text. Inst., 1997, 88, No.4, 465.
2. Cusick, G. E. J. Text. Inst., 1968, 59, 253.
3. D. H. Ballarel and C. M. Brown, Computer Vision, 1982. Prentice-Hall, Englewood Cliffs, NJ, USA.
4. AATCC Technical Mannual; 1985; 60,115.
5. S. Inui, A. Shibuya, and A Aisaka, J. Soc. Fib. Sci-tech. Japan, 1990, 46, No.6, 244.
6. G Stylios and J O Sotomi, J. Text. Inst., 1993, 84, No.4, 593.
7. D Marr in "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information", W H Freeman and Co., San Francisco, USA, 1982.
8. G K Stylios, Engineering of Textile Attributes, submitted for publication.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

A sewing rig with automatic feature extraction

Helder Carvalho[a], Fernando N.Ferreira[a], João L. Monteiro[b], Ana M. Rocha[a]

[a] Department of Textile Engineering, School of Engineering, University of Minho - Campus de Azurém, 4810 Guimarães, Portugal; email: hmc@eng.uminho.pt; fnunes@eng.uminho.pt; amrocha@eng.uminho.pt

[b] Department of Industrial Electronics Engineering, School of Engineering, University of Minho - Campus de Azurém, 4800 Guimarães, Portugal; email: joao.monteiro@dei.uminho.pt

This paper presents the development of a sewing rig that enables textile researchers to acquire, display and analyse data related to several operation parameters of an industrial sewing machine. The objective of this research is to quantify the sewing process, so that the influence of machine settings related to materials', needles' and threads' characteristics can be studied in-depth, and quality and machine set-up time issues can be adequately addressed. An overview of the sewing parameters involved, and of the acquisition and analysis system will be presented. The processing of signals related to needle penetration force into fabrics will be described, and test results will be displayed. Some particular properties of the measuring system based on piezoelectric sensors will be highlighted.

It will be shown that the equipment can be used as a sewability tester to determine needle penetration forces with different types of needles or materials, empowering apparel and textile manufacturers to tune their processes, avoiding quality problems. Some work on automatic detection of sewing defects that arise from needle penetration will be examined.

1.Introduction

Although at first sight apparel manufacturing appears to be a simple industrial process, a deeper observation shows that it is not. The transformation of 2-D shaped components into a 3-D garment by sewing is geometrically complex, taking into account that textile materials are flexible and have *drape*. Furthermore, sewing consists of passing threads through these flexible materials, materials that are available in a great variety of structures, with a great number of characteristics which influence the process.

Setting the machines correctly and choosing the right needles and threads is of utmost importance in achieving efficient, quality seams. Traditionally, this task is performed on an empirical basis. It is nowadays important to investigate the sewing process.

The objectives of this research are the following: 1. To gather knowledge about the sewing process and all the variables involved, so that the theoretical guidelines for achieving secure and high-quality seams can be strengthened; 2. To develop instruments and techniques that enable apparel and textile manufacturers to manage their processes so that quality problems are reduced or eliminated; 3. In the long-term, to develop add-on kits for sewing machines that will monitor their operation, detecting fault situations, and even adjust some parameters automatically.

Many works have been published regarding this subject. The system described in this paper has been presented earlier [1][2].

Signal processing algorithms have been improved, and many other functions have been added, which made possible to obtain some interesting results and interpretations of the sewing process.

Analyses of needle-penetration force, compression force on presser-foot [3] and thread tensions [4] is being continued. The system has been expanded to measure presser-foot displacement and thread consumption.

This paper will focus on an important aspect of sewing, the interaction between the needle and the fabric. Problems due to wrong needle selection or difficult materials arise very frequently in all branches of apparel manufacturing, and are also important in technical applications, for instance the manufacturing of airbags, sails or protective clothing.

2. Sewing parameters and their importance

The measurement system described was built on a three-thread overlock machine. The sewing parameters measured on this machine are of similar importance on other types of machines, and the insight gained by this work will provide the means of performing the same analysis in other cases in an identical way.

2.1. Thread tensions and consumption

Many different types of stitches exist for industrial use, using one or more threads, each having its own properties and applications. A stitch is geometrically well defined. If this geometry is kept, the stitch will perform adequately. A proper relation of thread consumption between the different threads in a stitch, achieved by correct thread tension adjustment, will guarantee a quality seam.

Thread tensions should also be matched to the material been sewn and to the function the seam has to fulfil. A finishing operation, for example, uses loose thread tensions, but a joining seam uses a stronger tension, so that the seam doesn't open up under stress (*seam grinning*).

Fine materials tend to *pucker* (becoming undulated) under high tensions, whereas heavy materials need higher tensions to hold the seam together.

2.2. Force on and displacement of the presser-foot

The *presser-foot* compresses the fabric being sewn against the *feed-dog*. The elliptical movement of the feed-dog provides the feeding of the material, and the amplitude of this movement determines *stitch length.*

Situations of mechanical resonance of the feeding systems causes irregular stitch length, and can be detected by dynamic measurement of presser-foot displacement and compression force. Evaluation of these parameters will enable a correct adjustment of static presser-foot force, and a monitoring of resonance situations is possible.

2.3. Needle penetration force

To form a stitch, the needle has to penetrate the fabric, and this penetration has to occur without damage. Two situations are to be avoided:

1. Yarn breaks - If the needle breaks the yarns, the fabric's structure is damaged at this point, causing the seam to fail when it is stressed, and resulting in a fault that cannot be repaired. This is particularly serious on safety seams, like in protective clothing or airbags, but is in general an unacceptable situation in quality articles.
2. Structural deformations - To penetrate the fabric, the needle has to create some space, distorting the structure of the fabric. If the fabric is unable to recover from this it might pucker, lowering the quality of the seam.

An adequate method to measure and characterise needle penetration will provide tools to minimise the effects described.

3. Overview of the measurement system

3.1.Sensors

The measurement of thread tensions is achieved by the use of cantilever beams placed into the thread paths that will acquire strain when stressed directly by the thread. This strain is proportional to thread tension, and is picked up by high-sensitivity semiconductor strain gauges. The sensors were designed to have a very high resonant frequency (about 10 KHz), being able to dynamically measure and track quick variations of the thread tensions.

Piezoelectric devices built directly into the needle-bar and the presser-foot spring sense the forces on needle-bar (which will be used to measure the needle penetration forces), and on the presser-foot. This aspect will be examined further on in this paper.

Presser-foot displacement is measured using an LVDT optimised for dynamic sensing. This method has shown to be very accurate and reliable even on high sewing speeds.

To obtain information on thread consumption high-resolution encoders, driven by the thread movement, have been introduced into the thread paths. Counting the pulses during a seam provides the mean to calculate consumption values for each thread.

3.2. Signal conditioning and acquisition

Signal conditioning/pre-processing boards were designed for each of the different types of measurement devices used.

The analogue signal conditioning boards can be externally programmed and have specific functions for sensor calibration.

All of the boards connect to a single bus, being connected to a data acquisition board in a PC. The data acquisition board used is a National Instruments LAB-PC+, which is a simple and economic board, but possesses all the functions necessary for the current needs of the system. In a future stage a board with a higher sample rate will be introduced.

The bus is based on the LAB-PC+'s connector, providing analogue and digital I/O, among other timing and triggering functions.

Signal acquisitions are made at a fixed sample frequency and constant sewing speed. In future, the system will be modified to acquire a fixed number of samples per stitch independently of sewing speed.

3.3. Software

The software was developed in National Instruments LABVIEW, an environment designed specifically for instrumentation applications. This made it possible to quickly develop an application with all the basic acquisition, file I/O, sensor calibration and hardware control functions.

Great advantages of LABVIEW are its powerful graphic display and signal processing tools. These were enhanced and combined with algorithms and functionalities designed specifically for the use in mind, resulting in a powerful application which provides all the tools necessary to an in-depth study of the signals acquired.

4.Analysis of needle-bar force signals

4.1.Set-up and properties of the piezoelectric sensor

Force on the needle-bar is measured to obtain information on the forces developed during fabric penetration and withdrawal using a piezoelectric sensor placed into the needle-bar system. However, with this arrangement two other variables are captured by the sensor: Needle-bar acceleration forces and forces produced by the thread movement through the needle *eye*. To examine the interaction between needle and fabric the latter can be eliminated by unthreading the machine. In a real-time situation, however, the effects have to be separated. Acceleration forces will always be present and have to be eliminated.

The piezoelectric sensor was chosen for its good dynamic capabilities in comparison to other types of devices. It has the major disadvantage of not being able to measure static values. The conditioning circuitry has a time constant associated, which is dependant on the values of the passive components in the first stage of signal conditioning - the *charge amplifier*. This time constant was adjusted to be sufficiently long to enable the circuit to measure the signal accurately at low sewing speeds, but is not sufficient to capture static values or very low-frequency components. As will be shown, the signal acquired at the needle-bar has fluctuations of its mean value due to the great variability of needle penetration force, fluctuation that is captured in a distorted way with the existing time constant, which will imply some corrections.

4.2. Needle-bar force signals

Figure 1 shows the force on the needle-bar generated during two complete stitch cycles with the machine operating unthreaded, with a medium-weight denim fabric (solid) and without fabric (dashed) at 2050 stitches per minute (spm), allowing a qualitative analysis of the signals.

The effect of needle penetration is clearly visible in peaks 1 and 2. Peak 1 corresponds to a force generated at the first contact of the needle tip with the fabric. Peak 2 is generated when penetration occurs and the eye and blade of the needle passes through the fabric structure.

As can be seen in figure 1, the signals present a slight shift of their dc value. This occurs due to the positive average value that is added to the signal when needle penetration force exists. The inability of the piezoelectric system to maintain this static or slowly changing value produces a fluctuation of the mean value of the signal. Furthermore, penetration forces are highly variable, meaning that the average value is different from stitch to stitch. This results in a fluctuation of the signal, and will have to be corrected. Peak 3 is the effect of needle withdrawal, an inverted peak, generally with a lower magnitude than 1 and 2.

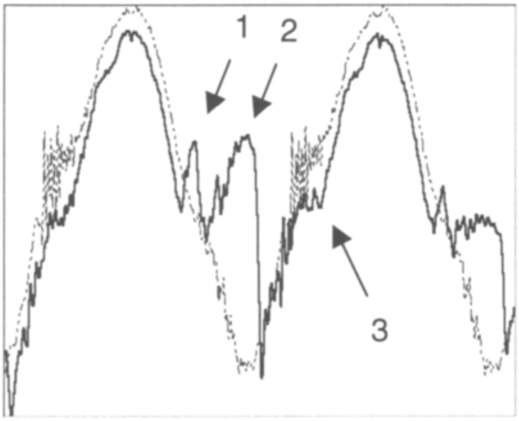


Figure 1: Force on needle-bar at 2050 spm, with medium-weight denim (solid) and without fabric (dashed)

4.3. Processing of the signals

Considering that the objective of this measurement is to analyse the effects of needle penetration, the need to eliminate needle-bar acceleration forces in the signals arises.

Spectral analysis has shown that harmonic components of the acceleration and needle penetration intersect, so that digital filtering or calculation in the frequency domain is a task that has to be approached with prudence.

A simple solution to obtaining the desired result consists is subtracting the pure acceleration signal (reference signal) from the signal which contains the penetration component. This approach is, in principle, not very reliable, because small speed variations can cause errors, and the acceleration component is likely to change in the presence of fabric. Nevertheless, the results obtained are interesting, and this technique serves as a good starting point to develop alternative signal processing algorithms that can improve reliability of the results and further automate the extraction of the desired values.

Subtraction of the signals in figure 1 leads to the result shown in figure 2. The same two stitches are shown, and the two phases of needle penetration have been delimited. Much care has to be taken when performing this subtraction. Sewing speed on both signals has to be constant during the acquisition and has to be the same. If this condition is ensured, errors will be minimised.

After this operation a zero correction has to be done stitch by stitch to compensate for zero drift originated to the filtering of zero and low-frequency components by the piezoelectric conditioning circuit. The arrows in figure 2 point to the moment where the needle is about to touch the fabric - at this time needle penetration force should be zero.

Other methods have to be developed to improve this measurement, and that will also make it possible to be used for real-time monitoring. It has not been possible with this approach to extract information about peak 3, because at the moment of its occurrence the reference signal presents some noise due to mechanical vibration.

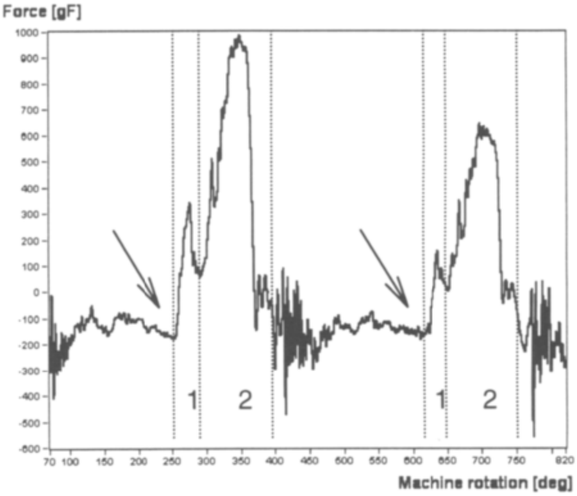


Figure2: Needle penetration signal after subtraction
Phase 1: First contact of needle tip
Phase 2: Penetration of eye and blade

A promising signal processing tool that will be considered is the decomposition of the signals using wavelets.

4.4.Calculations

The following values are calculated for each stitch and phase of penetration: the peak force value, the average, calculated as

$$\sum_{i=l}^{u} S(i)^2 \times \Delta t \tag{1}$$

and the energy , calculated as

$$\frac{1}{u-l+1} \times \sum_{i=l}^{u} S(i) \tag{2}$$

S(i): Signal value at sample i
l: Lower interval limit
u: Upper interval limit

5. Results

5.1.General results

All of the calculated values present a great variability. This behaviour was expected since the needle can enter the fabric through the yarn, gaps between yarns or yarn crossovers, resulting in penetration forces that are very different from each other. Some stitch cycles do not exhibit penetration force in phase 2.

As will be shown, differences caused by the use of different needles or materials are detectable, qualifying the system as a sewability tester. Different textile structures also produce different results, which points to the ability of the system to classify fabrics.

Frequency response of the measurement system has to be studied with more detail, since the values of peak, mean and energy of phase 1 tend to lower with increasing speed, whereas this increase is not as obvious for the same values in phase 2. It is unclear if this result is correct or if it is due to a loss in frequency response of the sensor or of the mechanical needle-bar/sensor assembly. The most significant results are presented in the next sections.

5.2. Difference between needles

The following graphs show the values of energy of phase 1 using a 70 and a 90 needle (0.7 and 0.9 mm diameter respectively)

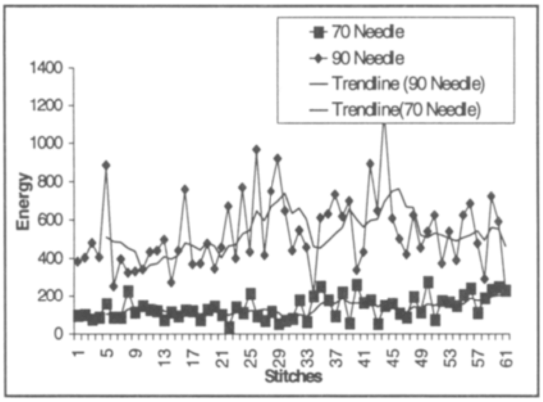


Figure 3: Difference of energy in phase 1 between 90 and 70 needle on a twill fabric at 2500 spm

Although a great variability of values exists, a clear difference between the two needles has been detected in all tests.

This difference has not been quite as clear on phase 2 of needle penetration, but the average of the values of phase 2 energy at several speeds and materials always expresses the distinction of the two needles, as shown in figure 4.

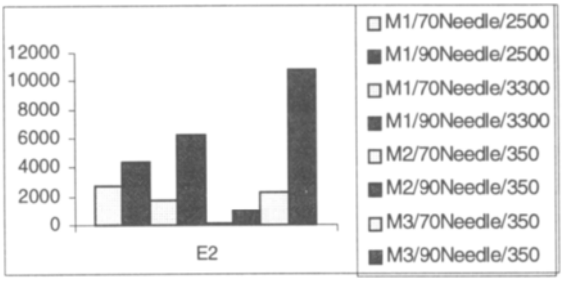
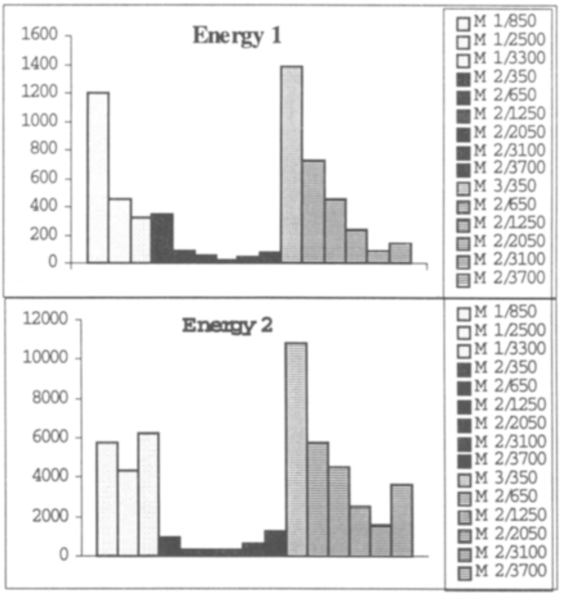


Figure 4: Average of Energy in Phase 2 for 70 and 90 needle, Materials 1,2 and 3 (Medium weight denim, light weight plain weave, medium weight twill) at different sewing speeds

5.3.Difference between sewing speeds and materials

The following graphs show the values of energy in phase 1 and 2 for the same three different materials and varying speeds, all with a 90 needle.



Figures 5 and 6: Energy of phase 1 and 2 with the three materials and varying speed. 90 needle used in all tests.

As expected, an obvious difference can be observed between the three different materials. Differences between sewing speeds show that the frequency response of the system has to be further studied, but comparisons at equal speeds are valid.

5.4. Fabric classification and automatic fault detection

An interesting way of observing the obtained results is by graphing energy 1 versus energy 2 values in an XY-Graph and observing the clustering of these pairs of values. It is to be expected that this clustering is dependent on the fabric structure. Furthermore, when yarn breaks or structural deformation of the fabric occur, a decrease of values of energy in phase 2 is foreseen, because the fabric structure is unable to resist to the needle's penetration.

Figures 7 and 8 show Energy1-Energy2 clustering for material 2 (light weight plain weave, regular structure) at 350 spm, where little seam damage was observed with both needles, and for material 3 (a medium weight denim, with a more complex structure), where a slight damage appeared with a 90 needle.

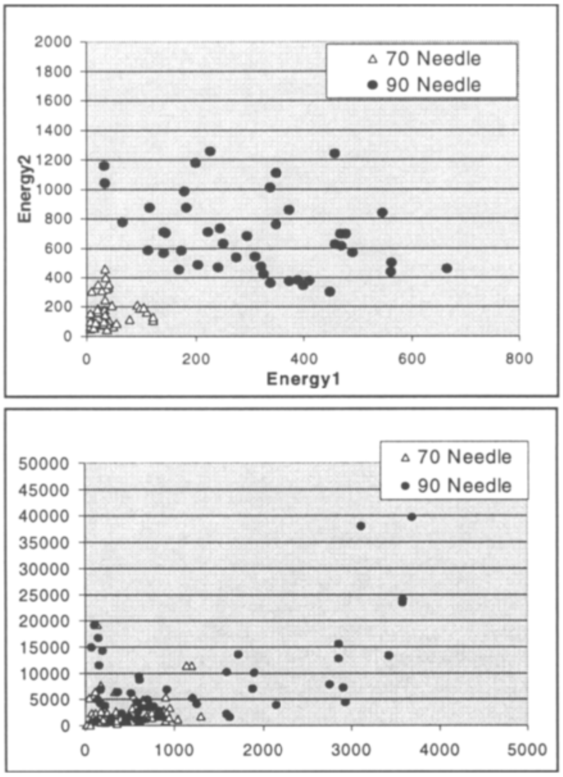


Figure 7 and 8: Clustering of Energy 1/Energy 2 pairs of values for Material 2 and 3. Sewing speed 350 spm, 70 and 90 needle

Clustering is more visible in the plain weave fabric than in the denim, mainly caused by the difference between the two structures, but it is believed that fabric damage will also influence clustering. This analysis is promising and will be extended to other practical cases. As withdrawal force, represented by peak 3, is certainly another indicator of material damage, it will be included in this analysis.

6.Conclusion

The equipment developed has shown to be able to be used as a sewability tester regarding needle penetration forces.

Signal processing has to be further developed to increase accuracy and ease of use, as well as to implement real-time monitoring of sewing conditions.

Frequency response issues have to be investigated so that the proper corrections can be introduced to compare sewing conditions at different speeds.

The clustering analysis presented seems to be a useful way to classify results, it may even be the clue for automatic fault detection and fabric classification.

References

1. H.Carvalho, F.Ferreira, J.Monteiro, "Measurements and Feature Extraction in High-Speed Sewing", *Proceedings of the ISIE'97*, pp 961-966;University of Minho, Guimarães, July 1997
2. H.Carvalho, "Medição e Análise de Parâmetros em Máquina de Costura Industrial", *Msc Thesis*; University of Minho, 1998
3. A.M.Rocha, "Contribution to the Automatic Control of Sewing Parameters: Study of Needle Penetration and Feeding Dynamics", *PhD Thesis*; University of Minho, Guimarães, 1996
4. M.Carvalho, F.Ferreira, "Study of Thread Tensions on an Overlock Sewing Machine", in *Proceedings of the 3rd International Conference on Mechatronics and Machine Vision in Practice*, pp 223-226; University of Minho, Guimarães, September 1996

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Study of the Feeding System of an Overlock Sewing Machine

Luís Ferreira da Silva(1), Mário Lima(1), Carlos Couto(2),
Fernando Ferreira(3), Daniel Andrade(2), Luís Pinto Ferreira(2)

(1) - Department of Mechanical Engineering
(2) - Department of Industrial Electronics
(3) - Department of Textile Engineering
School of Engineering, University of Minho • Campus de Azurém, 4800 Guimarães, Portugal

This paper describes the initial study of the research program being carried out on the feeding system of an overlock industrial sewing machine. This is expected to provide the basis for the development of a redesigned and optimised fabric feeding system, in which the role of an actuator integrated in the presser foot bar will be also discussed as a major contribution to achieving a desired dynamic behaviour at high sewing speeds.

1. INTRODUCTION

This paper is part of an important research work entitled "*Study of New Alternative Mechanisms for Controlling the Feeding Systems of Industrial Sewing Machines*". The main objectives of this project is to study, develop and integrate new actuation mechanisms in the presser foot bar to control the feeding system of an industrial sewing machine.

Therefore, this paper presents the study of the behaviour of the presser foot mechanism on an overlock sewing machine, specifying the technical characteristics of the most capable actuator for this application and the development of the control system.

1.1. The Problem

The feeding system of the overlock sewing machine used in this study is made up by three components: (1) a ***presser foot***, with an helical compression spring on the presser bar; (2) a ***throat plate*** and (3) a ***feed dog***, driven by a mechanism that provides differential feeding. These components are combined to achieve a controlled movement of the fabric and a correct stitch formation (figure 1).

However, some of the problems related with the sewing process rely on the interaction between the presser foot and the feed dog, since these two components do not effectively control the fabric. At high speed sewing the presser foot "bounces" and loses contact with the fabric, due to the fact that the feed dog rises above the throat plate and hits the presser foot. This lack of control usually results on an irregular stitch.

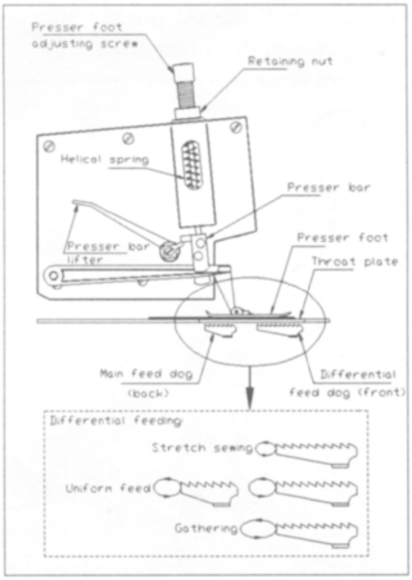


Fig. 1 - Feeding system of an overlock sewing machine.

2. THE APPROACH

To study the behaviour of the presser foot mechanism an advanced "sewability" tester was used, where the performance of needles, presser feet, feed dogs, fabrics and sewing threads can be assessed during high speed sewing. It was previously developed and improved by other research workers in our university [1-3]. They have instrumented a Singer overlock sewing machine, model 882U, with piezoelectric force transducers on the presser foot and needle bars, and semiconductor strain gauges on the thread paths, very close to the needle and loopers. A signal acquisition and analysis equipment was also developed for measuring the thread tensions, the presser foot bar compression force and the force on the needle bar. The study of these sewing parameters have led to a better understanding of the process and represents an important contribution to the on-line control of the sewing process.

3. EXPERIMENTAL WORK

A Sensotec LVDT was attached to the machine for real-time monitoring the movement of the presser bar [4], to study the problem of the presser foot during high speed sewing. This device together with a Kistler piezoelectric force transducer placed on the same bar will permit further understanding of the feeding system dynamics (figure 2).

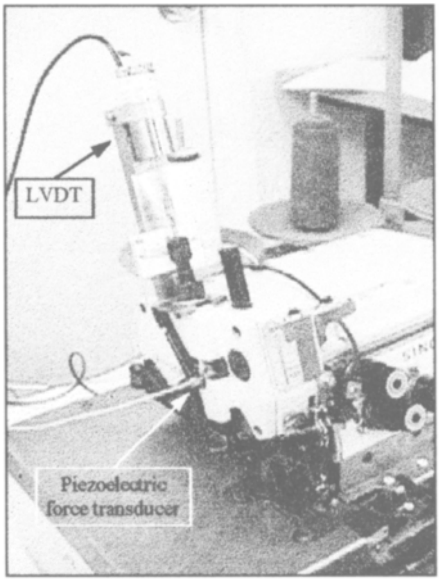


Fig. 2 - General view of the transducers arrangement.

In order to analyse the performance of the presser foot mechanism different experiments were made changing the machine speed. The displacement and compression forces of the presser foot bar, and the velocity and acceleration of this feeding component will also be presented and discussed in this paper.

The velocity and acceleration were obtained calculating the first and second derivative of the displacement-time curve of the presser bar. Before determining the derivative curves, a digital Butterworth filter (low-pass) was used for "cleaning" the sampled signal. The order and the cut-off frequency of the filter were chosen so that the resulting curve would be quite similar to the original one. The delay of the filter was also made as small as possible, so that the force compression waveforms could be easily compared with the obtained velocity and acceleration curves. The derivative curves were calculated using the following equation, which was simplified from the one used for the automatic peak detection of needle tip penetration [3]:

$$\frac{dy}{dt}(i) = \frac{1}{3}\cdot\sum_{j=0}^{2}\frac{y(i+j)-y(i+j-2)}{2\cdot\Delta t} \qquad (1)$$

where: y, signal value; i, sample i, in which is being calculated the derivative and Δt, the time between samples (reciprocal of the sample frequency).

4. RESULTS

In a stitch cycle, the feed dog is at the throat plate level, during its rising movement, at approximately 80 degrees, being its maximum level around 160 degrees. It will be at the throat plate level at approximately 260 degrees on its descending movement. During almost half period of the stitch cycle there is the fabric feeding and during the other half there is the stitch formation. Therefore, a stitch cycle corresponds to a 360 degrees' rotation of the sewing machine main shaft. The beginning of the cycle (0 degrees) is marked when the needle is on its lowest position, after the fabric penetration.

4.1. Displacement and Compression Force

The presser foot bar displacement and compression force waveforms for different machine speeds, 186, 2077, 2810 and 4654 stitches per minute (spm), are illustrated in figures 3 and 4. These experiments were made with no fabric, a presser foot pressure (or preloaded spring) of 20N, adjusted with the feed dog at its maximum level above the throat plate (which corresponds to 1mm) and with an uniform feed.

It can be observed that the presser bar displacement increases with the machine speed at the point corresponding to the maximum elevation of the feed dog above the throat plate (figure 3). At 2077 spm, however, a second displacement occurs, just after the feed dog is below the throat plate. This effect is quite notorious at 4654 spm, which indicates a second contact loss between the presser foot and the feed dog mechanisms.

The presser foot bar compression forces (relative values) also increase with the machine speed (figure 4). Around 80 degrees there is an increase of the compression forces, when the feed dog is at the throat plate level on its rising movement. During the feeding period, between 80 and 260 degrees, there is a wide variation of the forces involved, which means that the spring forces are not the most adequate to guarantee contact between the two mechanisms, and a contact loss happens. Just after the feed dog has been lowered beneath the throat plate (after 260 degrees), a larger peak force was observed. This means that the presser foot re-establish contact with the throat plate. This significant peak force increases with the machine speed.

A second set of experiments were made. A jersey fabric (K=17, medium weight=159,1 g/m2) was used. Two plies were sewn with a stitch density of 4 stitches per cm. The other machine settings were identical to the ones used before. The machine speed varied between 2000, 3000 and 4760 spm, and about 100 stitches were read and stored for each tested sample.

It can also be seen that the displacements of the presser bar increases with the machine speed (figure 5). The contact loss between the presser foot and the feed dog also occurs when the feed dog is completely above the throat plate. A second contact loss is once again observed at the (maximum) speed of 4760 spm, after 260 degrees.

The compression force waveforms show, as mentioned, an increase after 80 degrees, when the feed dog rises above the throat plate (figure 6). During the feeding period, a variation of the presser foot bar compression forces up to 260 degrees was observed once more. At this point, or soon after, the final peak forces are also significant. However, it must be regarded that the magnitude of these compression forces are now quite smaller than the ones already discussed (see figure 4).

4.2. Kinematics

The presser bar graphic kinematics (including the displacement, velocity and acceleration) are illustrated in figures 7 to 9.

The calculated velocity of the presser bar reveals certain well defined peaks, during the rising and descending movements of the presser bar. The magnitude of these peaks increase with the machine speed.

It is possible to identify several positive and negative acceleration peaks by analysing the acceleration curves of the presser bar, during a stitch cycle. These negative peaks, that increase with the machine speed, are very important for identifying the "critical points" in a stitch cycle. But, just one of these peaks is more important than the others. This peak, the most negative of all, occurs between the maximum velocity point of the feed dog and the position corresponding to the feed dog at the top. It is here that a higher possibility for a contact loss may occur. Being the acceleration maximum negative, the inertia force will be maximum positive and opposes the weight and the spring forces. The condition for a contact loss can be written as:

$$F_i > W + F_s + P_{fp} \qquad (2)$$

where: F_i, is the inertia force; W, the weight of the presser foot mechanism; F_s, the spring force and P_{fp}, the presser foot pressure (preloaded spring).

It can also be noted that, at the speed of 4760 spm, a second negative acceleration peak is surely responsible for a second displacement. This explains the contact losses observed in fig. 3 and 5.

5. CONTRIBUTION FOR A REDESIGNED FEEDING SYSTEM

New alternative actuation mechanisms to control the feeding system of an overlock sewing machine are being studied, in order to achieve the desired dynamic behaviour of the presser foot bar at high speeds. The selected actuator will allow the development of an actuation's efficient control system, in real-time, on the presser foot bar. Thus, an actuator capable to function in a "continuous" mode, that presents a linearity between the force applied and the displacements intended, constitutes, surely, a good solution.

Today, the most used actuation mechanisms are the piezoelectrics, magnetostrictives, shape memory alloys and electromagnetics actuators [5].

Piezoelectric actuators are excellent for several purposes because of their quick response, small size, low power consumption and high power density. However, hysteresis and creep, both in the order of 15-20%, applied voltages in the region of 1-2kV and small output movements, typically 15μm, are not only well-known limitations but also strong reasons not to opt for this actuator.

Magnetostrictive actuators are used in low-frequency applications and like piezoelectric actuators do not accommodate large deflections (<250μm). However, they require significantly lower voltage and the desired performance is lost at lower and higher temperatures.

The shape memory alloy technology requires heat as a primary driving mechanism. But, the main drawback of these actuators is their very slow rate of change because actuation depends on heating and cooling: their response depends upon the velocity of temperature variations. The dynamic response of this device is almost invariably very slow too [6].

Finally, another method of actuation uses the electromagnetic mechanism. The primary advantages of this technology are its quick potential actuation and lower power consumption. One electromagnetic device is the solenoid, that consists of a coil for generating an electromagnetic field; a plunger acted upon by the field to convert it into useful mechanical work, and an enclosing structure to concentrate the lines of the flux generated [7].

Among all kinds of solenoids, there is one that may be useful for our purpose. The proportional force solenoids offer a stroke (impulse/force) proportional to the injected current. This feature will control automatically the position of the solenoids' plunger by varying the current injected in. With this particular solenoid, it is possible to obtain the force and displacement outputs necessary for our application need. However, this subject is still being investigated according to what was presented and discussed in the previous section.

6. CONCLUSIONS AND FUTURE WORK

The first sewing machines could stitch at 200-300 spm, but, nowadays, it is even possible to work at speeds of approximately 9000 spm. For that very reason, the performance of the presser foot mechanism is not the most effective, as it was once thought.

Although more experiments to properly study the presser foot mechanism and the influence of other machine parameters on its performance are needed, we can conclude from the available results that the feed dog and the presser foot mechanisms lose contact when machine speed increases, with or without the fabric. In the early 70's, Johnson [8] discussed in detail this "bouncing" phenomenon.

During the feeding period, the compression force waveforms also show a wide variation, which indicates that the performance of the helical spring is not the most adequate for guaranteeing the contact between the two feeding components.

The "critical points" in the stitching cycle, where the feed dog and the presser foot are about to lose contact or, in fact, are not in physical contact, which corresponds to the most negative values of acceleration, are identified according to the waveforms of the velocity and acceleration of the presser bar, in conjunction with the other two waveforms.

These results are an important instrument for the evaluation and prediction of materials behaviour in sewing and will also provide the basis for future work, namely developing a redesigned fabric feeding system with a controlled actuation and improving fabric control to achieve better sewing quality at higher speeds.

REFERENCES

[1] Rocha, A. M.; Ferreira, F. N.; Araújo, M.; Monteiro, J.; Couto, C.; Lima, M. F.; *Mechatronics in Apparel: Control, Management and Innovation on the Sewing Process*, Proceedings of the Mechatronics'96 Conference, Vol. II, pp. 109-114, Guimarães, September, 1996.

[2] Carvalho, M.; Ferreira, F. N., *Study of Thread Tensions in an Overlock Sewing Machine*, ViCAM - Vision and Control Aspects of Mechatronics, pp. 223-226, Guimarães, September, 1996.

[3] Carvalho, H.; Monteiro, J.; Ferreira, F. N., *Measurements and Feature Extraction in High-Speed Sewing*, Proceedings of The IEEE: ISIE'97 - International Symposium on Industrial Electronics, Vol. 3, pp. 961-966, University of Minho, Guimarães, July, 1997.

[4] Silva, Luís F.; Lima, M.; Couto, C.; Ferreira, F. N.; Rocha, A. M.; Carvalho, H., *Estudo do Mecanismo do Calcador de uma Máquina de Costura Corta-e-Cose*, Jornadas de Engenharia Mecânica, Revista FIMEC'98, pp. 24-26, Universidade do Minho, Maio, 1998.

[5] Dorey, A. P.; Moore, J. H. (eds.), *Advances in Actuators*, Institute of Physics Publishing, Bristol and Philadelphia, 1995.

[6] Gilbertson, R. G.; Busch, J. D., *A Survey of Micro-Actuator Technologies for Future Spacecraft Missions*, presented at the conference Practical Robotic Interstellar Flight: Are We Ready?, New York, 1994.

[7] Charkey, E. S., Electromechanical System Components, John Wiley & Sons, 1972.

[8] Johnson, E. M., *Some Factors Affecting the Performance of High Speed Sewing Machines*, Clothing Research Journal, Vol. 1, Nº. 1, pp. 3-35, 1973.

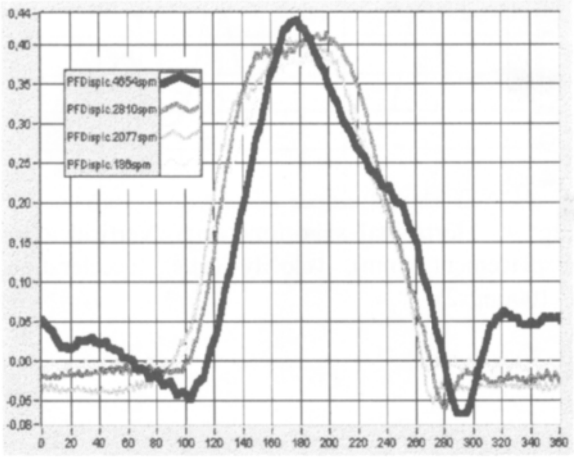


Fig. 3 - Horizontal axis: stitch cycle (degrees); Vertical axis: presser bar displacement (mm).

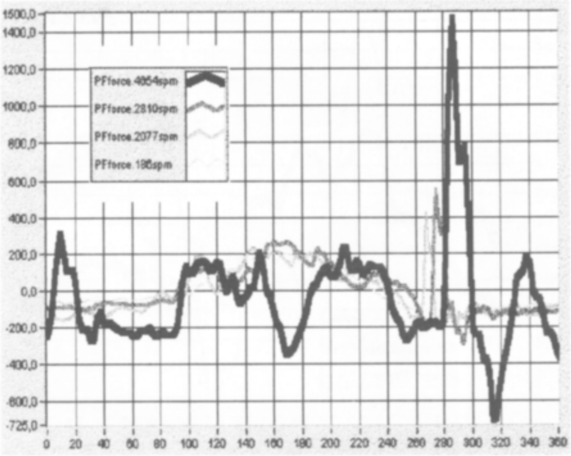


Fig. 4 - Horizontal axis: stitch cycle (degrees); Vertical axis: presser foot bar compression force, relative values (gf).

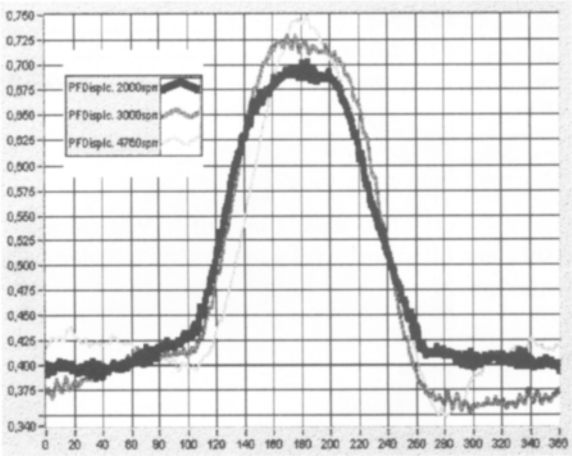


Fig. 5 - Horizontal axis: stitch cycle (degrees); Vertical axis: presser bar displacement (mm).

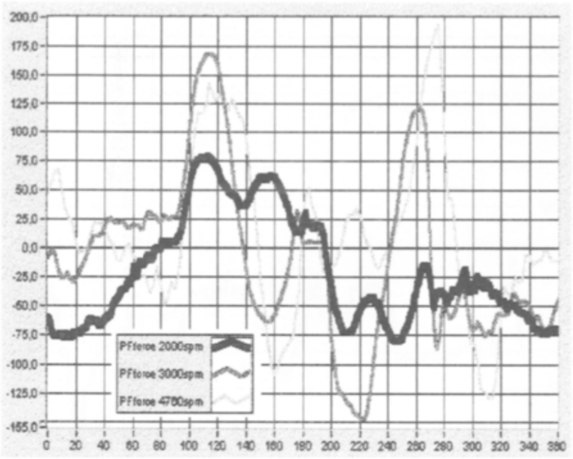


Fig. 6 - Horizontal axis: stitch cycle (degrees); Vertical axis: presser foot bar compression force, relative values (gf).

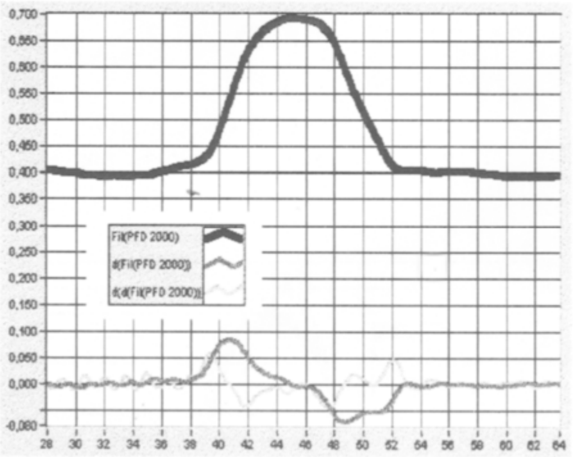


Fig. 7 - Horizontal axis: time (ms); Vertical axis: displacement (mm), velocity (m/s), acceleration (×1000m/s2).

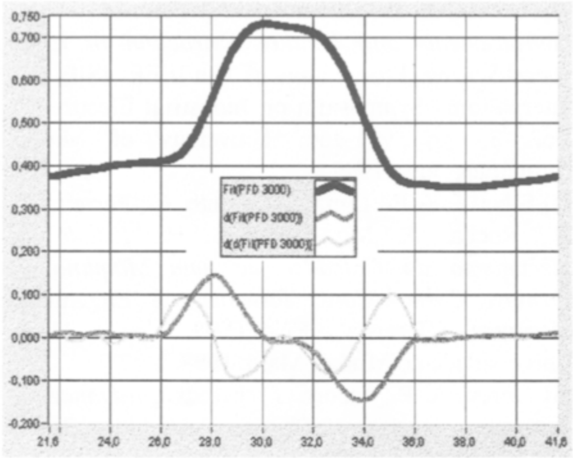


Fig. 8 - Horizontal axis: time (ms); Vertical axis: displacement (mm), velocity (m/s), acceleration (×1000m/s2).

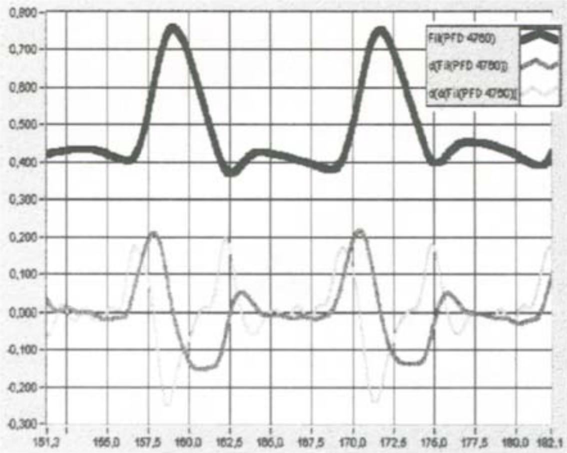


Fig. 9 - Horizontal axis: time (ms); Vertical axis: displacement (mm), velocity (m/s), acceleration (×1000m/s2).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

Determining the effectiveness of computer vision when measuring yarn interlace at high speeds.

M.P. Millman, M. Acar, M.R. Jackson

Department of Mechanical Engineering
Loughborough University
Loughborough, Leicestershire
LE11 3TU, UK

The aim of the work of which this paper is a part, is to determine the effectiveness of measuring many characteristics of different types of yarn using computer vision. This paper deals specifically with interlace of false twist textured yarns. A system has been built which is capable of analysing yarns at slow speed / high resolution, and at high speeds / low resolution. An outline description of the system is given here, along with descriptions of techniques used in the hardware and software which substantially improved the system performace. To monitor interlace, a recently developed signal processing algorithm is explained which enables reliable interlace detection, and is effective against noise.

Several tests are carried out, which prove the ability of the system to match the results from careful manual inspection of the yarn. The final test is to run the yarn at a high speed, and to compare the results with the slow speed analysis. The results show that at high speeds, the system is reliable. A discussion of some hardware requirements for an on-line optimised commercial system is given.

1. INTRODUCTION

Computer vision is becoming more affordable as both a research and in-process quality control tool. It has the advantage of providing increased amounts of characterisation of the test product, at high speeds and resolutions, as well as being non-contact.

Synthetic continuous filament yarns are usually textured prior to their successive usage in further processes, such as weaving and knitting [1]. The purpose of conventional texturing is to impart texture to the regularly packed, flat filament yarn to give comfort and aesthetics to the finished product. The most common texturing technique is the false-twist texturing process which produces 'stretch' yarns that are highly extensible even under very low loads. This arises from heat setting the twisted deformations introduced to the yarn during the texturing process.

False-twist textured yarns lack inter-filament cohesion, and consequently a number of difficulties are observed during unwinding and fabric forming processes. One of the modern techniques of imparting such cohesion is air-intermingling of false-twist textured yarns. Intermingling is a process which deploys a nozzle to create a very turbulent high speed air flow. This imparts regular but intermittently entangled nodes to the open structure of the textured yarn which are commonly known in industry as 'nips' [2].

The frequency and regularity of the nips together with their stability under applied loads are important criteria in the assessment of the performance of mingling nozzles and the quality of the intermingled yarns.

2. SYSTEM OUTLINE

The overall system is divided into two main parts; the yarn transport and the yarn imaging/analysis. For each part, a separate PC computer is used, and these are connected together. The imaging PC controls all operations, and sends instructions to the transport control PC, which in turn can operate in a stand alone fashion if necessary.

2.1 Yarn imaging

The yarn is imaged using a 1024 pixel line scan CCD camera which scans up to 10 kHz. For a yarn travelling at 10 m/s the scan to scan resolution is 1 mm. Two orthogonal views of the yarn in the same cross sectional plane are used to characterise the yarn. A mirror inclined at 45° to the viewing plane achieves this as shown in Figure 1. This technique has been shown to provide more accurate evenness data for staple fibre yarns [3]. The magnification is set to 1, giving a pixel resolution of 13 microns.

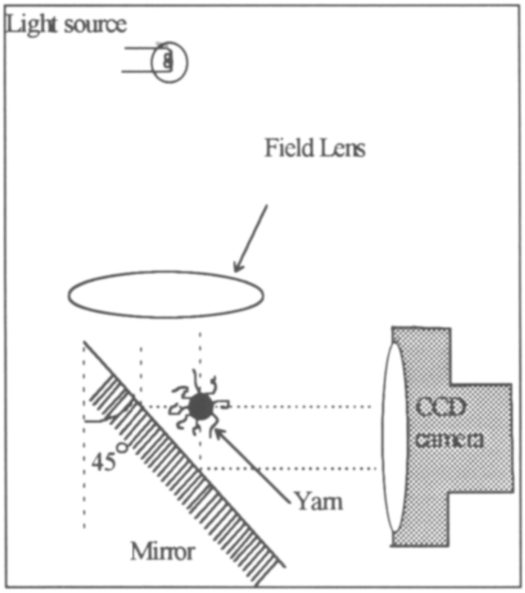


Figure 1: Arrangement for obtaining two orthogonal views of yarn image

A key optical technique when imaging yarns for computer vision is to use a field lens before the yarn. This effectively focuses the light source (30 watt halogen bulb) onto the aperture of the camera, which causes a much larger degree of light intensity to be seen at the CCD. Furthermore, the light intensity profile across the CCD is much more constant than without a field lens and this is critical when attempting to image loose fibres/broken filaments (Figure 2a & b). The alternative to this is to use a diffuser instead of the field lens, but this absorbs a significant proportion of light, and does not give as even an illumination.

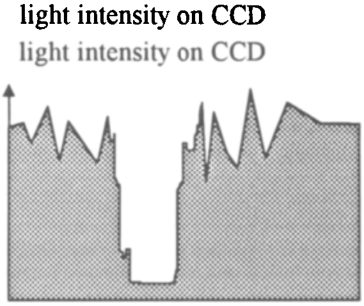


Figure 2a - Intensity profile of a yarn image without using a field lens or diffuser.

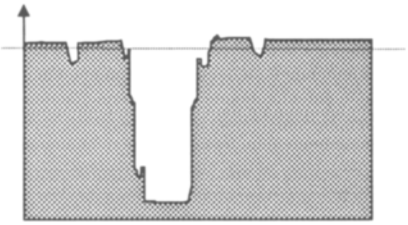


Figure 2b - Intensity profile of a yarn image using a field lens or diffuser. The dotted line represents a threshold used for detecting broken filaments/loose fibres.

The camera is interfaced to a PC using an Optimum Vision OVB10 board. This binarises the incoming grey level analogue signal, and sends the edge data to the PC on a FIFO stack system. For nip detection, the grey level data is not required, and binarising the image is all that is required from an imaging point of view. This isn't the case for yarn hairiness or evenness analysis [5]

An important part of the software is the acquisition of the edge data from the OVB10, and is written in C code. However, the part which reads data from the OVB10 board had to be optimised to the full, in order to be quick enough to receive the data. An important and very powerful trick was utilised here; The original C code, assembler output was taken, and modified by using most of the Intel 80386 register set, with careful thought given to instruction timings etc. The result is faster code, and the image acquisition is less prone to saturation if there is too much edge data from the OVB10. This technique can substantially improve the performance of many software based real time applications, where the code is written in a high/medium level language.

2.2 Low Speed Yarn Transport

An off-line computer controlled yarn transport has been developed which can transport the yarn at slow speeds (7 cm/s), but with highly responsive start / stop properties using stepper motors. The tension of intermingled yarns affects the nip dimensions as the yarns stretch. Therefore a closed loop PID tension control was built, such that the tension remained constant, and could be held steady at low tensions (~5cN). The tension data is also monitored, and sent to the imaging PC at the end of a run to check if there were no glitches in the yarn transport.

The yarn is engaged to the stepper motors by means of solenoid actuated capstan rollers, which pinch the yarn against the motor shaft. The two motors are at either end of the imaging area. The motor situated after the imaging area is run at constant speed, whereas the motor before the area is controlled by the tension controller software.

2.3 High speed yarn transport

A high speed yarn transport (up to 10 m/s) was made such that yarns could be tested at high production speeds. The results could be compared to those at lower speeds to show the effect of reducing the scan to scan resolution .

Two printed circuit DC servo motors were used in the same fashion as the stepper motors above i.e. with one after and one before the imaging area. The motor shafts have 50 mm diameter rollers fixed, and the yarn is wound several times using special guides onto these rollers. To enable a positive feed, each motor roller has a rubber roller pressed against it to hold the yarn. The motors have encoder wheels attached, signals from which are fed into a dual channel pulse counter. This enables the speed of the motor speeds to be monitored, as well as their ratio (percentage extension of the yarn).

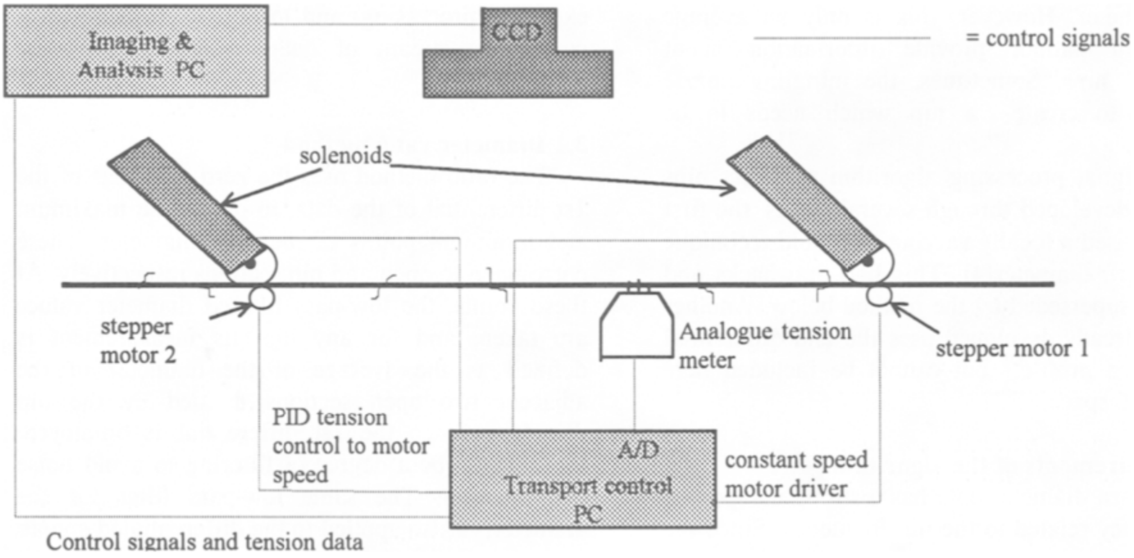


Figure 3 - Overview of the slow speed / high resolution system.

Careful consideration had to be given to yarn guiding near the measuring area. Any suspended length of yarn will tend to vibrate. If the yarn section being imaged is vibrating, the effect will be that the yarn image has a periodical increase and decrease in width. This is because a lateral velocity makes the yarn appear over a wider area for a finite integration time of the CCD scan. To prevent this, two guides were placed as near to the measuring area as possible without affecting the light intensity.

2.4 Manual yarn inspection

In order to verify the validity of the results from the computer system, it was necessary to use a microscope, and to manually record the nips. A sample section would be suspended across an inspection table along with a measure, and an appropriate tension was added using a weight at one end. A microscope which could slide from one end to the other was used to check the surface structure of questionable nips. Most nips could be observed with the naked eye, but some needed the yarn tension to be released to check the nip integrity

3. NIP MEASUREMENT METHOD

The average nip frequency (and hence average nip pitch) can approximately be determined from the fundamental peak in the frequency spectrum of yarn diameter. However, this is only an average value and cannot provide information about individual nips. Sometimes, the mingling nozzle may fail to create a nip which needs to be identified.

The signal processing algorithm to detect nips has been developed through several stages, the first of which used a locally varying threshold technique on the yarn diameter[4]. This had drawbacks and has been superseded by the method below. Another method already developed uses the 2nd differential of the yarn profile, but cannot be included here because of space.

3.1 Requirements of the algorithm

The yarn diameter data has a strong component of frequency related to the nip frequency. Since the nip distances can vary considerably, this frequency content of the data is spread over a region of the discrete fourier spectrum. Any filtering of the data must not attenuate these frequencies.

Signal noise can be in the form of oscillations caused by yarn vibration, or rapid changes in the yarn diameter due to surface filaments. This region is higher in frequency than the nip frequency range, but can overlap, especially when vibration is present.

Low frequency noise is caused when the apparent diameter of the yarn changes in both the nip and open section areas together over a wider distance causing problems to any form of thresholding.

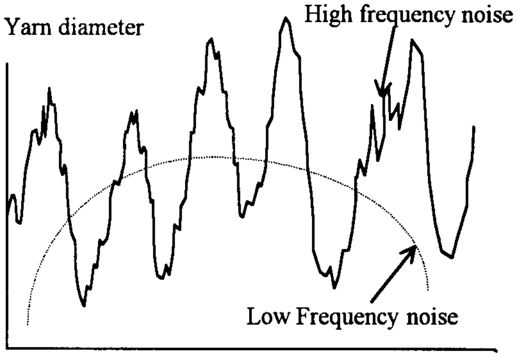


Figure 4 - The presence of noise on the data.

The algorithm needs to avoid requiring excessive processing, and to be able to work on a continuous stream of data for possible on-line inspection.

3.1 Diameter ratio method

The ratio method uses the zero crossings of the 1st differential of the data, to determine maximum and minimum points of the yarn diameter. These correspond to open and nip sections respectively. At these points, the low-pass filtered diameter values are taken, and for any nip, its measurement is defined as the average of the diameter of the adjacent two open sections divided by the nip diameter. Since the 1st differential is employed, there has to be a degree of filtering to avoid noise amplification. The same low-pass filter for the diameter, is also applied to the differentiated signal. Experimentation found that a cut off frequency of 3 times the average nip frequency is ideal.

The ratio measurement is unit-less, independent of nip length, not sensitive to low frequency variations in diameter, not sensitive to noise and is not computationally demanding.

4. FILTER DEVELOPMENT AND HARDWARE CONSIDERATIONS

The ratio method requires a digital filter which differentiates the signal, and low-passes the result. The number of coefficients of the filter needs to be kept to a minimum such that a cheaper microprocessor could be capable of the task, since floating point operations are expensive in time.

It was observed that by convolving a moving average filter of ones [1/n 1/n 1/n 1/n 1/n ...] with a differentiator [-0.5,0.5], the result is a filter like this: [-0.5 0 0 0 0 ... +0.5]. Since there are only two non-zero coefficients, it means that this filter only requires two multiplication's per digit for any required cut-off frequency. Since it is an FIR filter, its lag (group delay) is constant across the frequency response, which is also desirable.

Although implemented on a PC at present, it is feasible to implement the processing side on a dedicated DSP or micro-controller. A 16 MHz PIC micro-controller would have 1600 machine cycles per scan to process line-scan data at 10 KHz. This is sufficient time to process the data using the ratio method.

5. THE TESTS AND DATA ANALYSIS

For the tests, a synthetic continuous multi-filament false twisted intermingled yarn was used.

5.1 Slow speed versus manual inspection

The first test compares the nip to nip distances observed using the microscope mount, to those obtained by the computer under slow speed scanning conditions. A 2m sample was first examined using manual inspection, and the positions of its nips were recorded, along with an indication of how well formed the nips appeared (extremely poor, very poor, poor, normal). The same sample was then examined using the slow speed/high resolution computer vision system with the diameter ratio method. The distances between each nip were then compared and are shown below in Figure 5.

5.2 Slow speed versus high speed long length tests

The second test scans a yarn length of 50m under slow speed scanning conditions, and compares the results with those obtained using the high speed transport at production speeds, but using the same threshold setting for the diameter method..

For the high speed test, the camera scan rate is set to 10 KHz, the yarn speed is 7.5 m/s, and the motor speed ratio (percentage slip) is 1.025. The mean nip count per metre and standard deviation for both speeds is shown in Table 1 for both tests.

5.3 Test results

It is clear from Figure 5 that that for the slow speed/manual inspection test, the correlation is exact, in that if any nip was missing in one set of data, but not in the other, then the lines on the graph would become out of step by 1 nip at that point, and this isn't observed here. It was also observed that the threshold level in the diameter ratio method could selectively exclude those nips labelled extremely poor or very poor, which if also excluded from the manual data, would give the same exact correlation. The test proves the system for this type of yarn, and also demonstrates the use of the threshold to selectively exclude nips of poor quality.

For the second test, the mean nip count per metre and the standard deviation for both speeds are very close. The standard deviation is less than 2 %, showing that either the same miscalculation is happening consistently or the detection is reliable.

More importantly, the number of nips per metre is the same for tests 2 and 3 as for test 1. This indicates that the system is working properly.

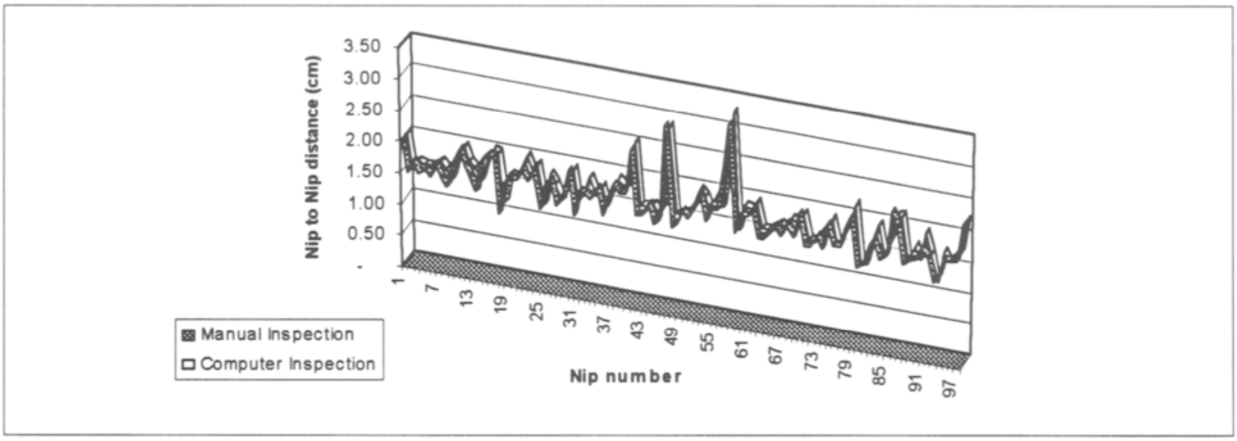


Figure 5 - Results from the slow speed /manual inspection test.

Table 1 - Tests results

Test	sample length (m)	yarn speed (cm/s)	scan resolution (mm/scan)	mean nips /m	Standard deviation (nips/m)
Manual	1.65	5.0	0.172	58.4	N/A
Low speed	50.0	5.0	0.172	56.9	1.58
High speed	50.0	7.5	0.78	54.2	1.64

6. CONCLUSIONS

The tests show that nip detection using machine vision is reliable. This is shown for both high and low yarn speed conditions. The system is shown to be technically capable of in-process inspection, since the speed used in the test 3 is similar to those used in production. The ratio method processing algorithm is shown to be effective at ignoring unwanted frequency components in the yarn profile, and is not computationally demanding.

REFERENCES

1. Wray G. R. and Acar M. Supersonic Jet Texturing of Yarns, Proceedings of the IMechE, Part B: *Journal of Engineering Manufacture*, 1990, Vol. **204**, 71-89.
2. Demir A. and Acar M. An Insight into the Intermingling Process, in: *Textiles: Fashioning the Future* , Manchester, 1989, 169-191.
3. Wegner E and Hedwig R, Forschungsberichte des Landes Nordrhein-Westfalen, No.2007,1968
4. Acar M, Jackson MR and Millman MP, High speed yarn characterisation and monitoring using a vision based scanning system, *Conf on Textile Process Control, 2001*, Manchester University UMIST, 1995.
5. Millman M.P, Acar M and Jackson M.R., Optical Resolution Limits and Digital Image Processing Techniques for Yarn Image Acquisition., *Mechatronics'96 - International Conference,* 1996, University of Minho, Vol. 2, 2.115 -2.122

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

In-Process Fabric Defect Detection and Identification

J. Lewis Dorrity[a] and George J. Vachtsevanos[b]

[a]School of Textile & Fiber Engineering, Georgia Institute of Technology
Atlanta, Georgia, USA 30332-0295

[b]School of Electrical & Computer Engineering, , Georgia Institute of Technology
Atlanta, Georgia, USA 30332-0250

High quality is the key to global competitiveness in most industries and certainly in the textile industry. This is most certainly true for weaving operations where the value added of this process makes mistakes very costly. Textile companies continue to look for ways to monitor and control quality at all of the processing stages. This paper discusses new technology developed at the Georgia Institute of Technology in Atlanta, Georgia, USA which detects defects during the weaving process thereby emphasizing defect prevention. Also discussed is the philosophy of automatic fabric inspection.

1. INTRODUCTION

The Textile industry in the U. S. A. and around the world is increasingly concerned with quality of their products. The global market pressures put higher and higher demands on quality. Machinery manufacturers feel the pressure from their customers to add sensing and control systems which either correct changes in operating conditions automatically or alert the operator to substandard conditions so that the machine may conditions may be corrected. This is certainly true for the weaving operations where the value added of this makes mistakes very costly. Computerized systems to insure high quality are in increasing demand. New technologies are being incorporated into new machines as the rapidly decreasing costs of computing power and sensing improve the economics of doing so.

Textile companies continue to look for ways to monitor and control quality at all of the processing stages. This is typified by the 100% quality monitoring of open-end spinning processes and sliver processing equipment in many enterprises. Quality monitoring of open-end spinning has become virtually required rather than optional equipment when purchasing decisions are made. The high speeds of production and inevitable deterioration in quality caused by contaminants building up in rotors render this necessary. The spinner cannot afford to allow the processing of off-quality yarn to continue because of its effect on the fabric it would ultimately cause to be substandard. The cost of the monitoring systems is considered as part of the cost of operating this type of spinning.

Many weaving operations are considering new technologies for monitoring the production of woven fabrics. Over the last 20 years, there have been a number of systems that have been marketed to automatically inspect fabrics for defects. Most have been for greige (undyed) fabrics that attempted to replace the visual inspection after the fabric was removed from the weaving machines. Such systems may save on labor costs, but require further handling to remove off-quality sections in another process. Typical of these is the Flying Spot Laser scanning unit marketed by Ford Aerospace in the late 1970s.

The Ford Aerospace system resulted from the technology transfer from a system designed to inspect the float glass process. This system employed somewhat oval shaped laser spots oriented in the warp and filling directions to enhance the response to line defects in the two directions.

Other systems were designed and patented such as those of Sick Optik and Recognition Systems, Inc. (RSI) [1]. The former device used a specially designed optical cylinder to collect signals from transmitted laser light through the fabric. The signal was collected and analyzed through analog means. The RSI system used a large spot (about four centimeters in diameter) and numerous mirrors to produce diffraction patterns that were collected on linear diode arrays in the warp and filling directions. Any distortion of the weave pattern would result in a distinct change in the diffraction pattern, which could be sensed by the digital electronics

Presently there is a system on the market by Elbit Vision Systems which inspects fabric at speeds of more than 100 meters per minute. These systems are primarily employed to replace the visual inspection of greige fabrics after the fabric roll is removed from the weaving machine. Some applications of this are also in-line with certain denim finishing processes.

A research project sponsored by the AMTEX Consortium developed a prototype vision-based system using a line-scan camera and high-speed processing to obtain and analyze images from a weaving machine. This part of the project has ceased, but was recently reported on at an IEEE conference in Charlotte, N. C., USA. [2]

More recently, at the Georgia Institute of Technology in Atlanta, Ga. , USA, the authors have developed vision-based technology along with wavelet analysis techniques to allow defect detection during the weaving process. This work has been sponsored by a grant from the National Textile Center (NTC) which is a consortium of universities having departments specializing in textile research. A U. S. patent of this technology has been granted. The system employs circuitry that integrates the image acquisition hardware with signal processing technology to create detector modules for inspection of fabric in process. The major components of this system are the detector board (CCD & DSP), controller board, and the user interface electronics. The licensee of the technology, Appalachian Electronic Instruments, Inc. is beginning commercialization. The details of this development are given below.

2. PHILOSOPHY OF AUTOMATIC INSPECTION

Automatic fabric inspection is employed in two basic ways. The most common way is to replace the process by which employees review fabric as it passes over an inspection table. Cloth is graded based on an industry standard or one which is agreed to by the customer. When such systems are employed, most plants would also employ persons to roam the weave room inspecting the cloth while still on the weaving machines. A typical weave room might have two or three inspectors per shift looking at something less than 100% of the production. This process may find many of the running warp defects, but the more elusive repetitive defects, such as broken picks, would be more difficult to track. To address this problem, some mills have installed terminals on the weaving machines to allow entry of defects by the weavers or roving inspectors. The economics in this scheme would be based primarily on labor savings in the inspection department.

Integrating the inspection process into the weaving process is the alternative. Such a system not only finds defects, but also stops the production of running or repetitive defects by alerting the operator to the condition. The emphasis in this scheme is on preventing the continuing production of defective fabric. When a roll of fabric is removed from the machine, the quality is known and the sections which need to be reviewed or removed are known. No roving inspectors are required and manual review after doffing should be minimized. The result is less off-quality production, less handling of fabric and lower labor costs. One requirement would be a greater emphasis on eliminating such things as hanging strings by the weaver so that burling is not required. Then, fabric rolls that are judged first quality may be shipped without rehandling and those which require sections to be removed could be handled much more efficiently.

3. SYSTEM DEVELOPED AT GEORGIA TECH

Beginning in 1994, researchers at Georgia Tech began to develop an affordable system to find defects while in production in weaving. There are many requirements and constraints for such a system including accuracy, robustness and low cost. This system is based on a CCD array camera and DSP processors to analyze data from the cameras.

3.1 System Overview

The automated fabric inspection system is composed of two major components: An off-line

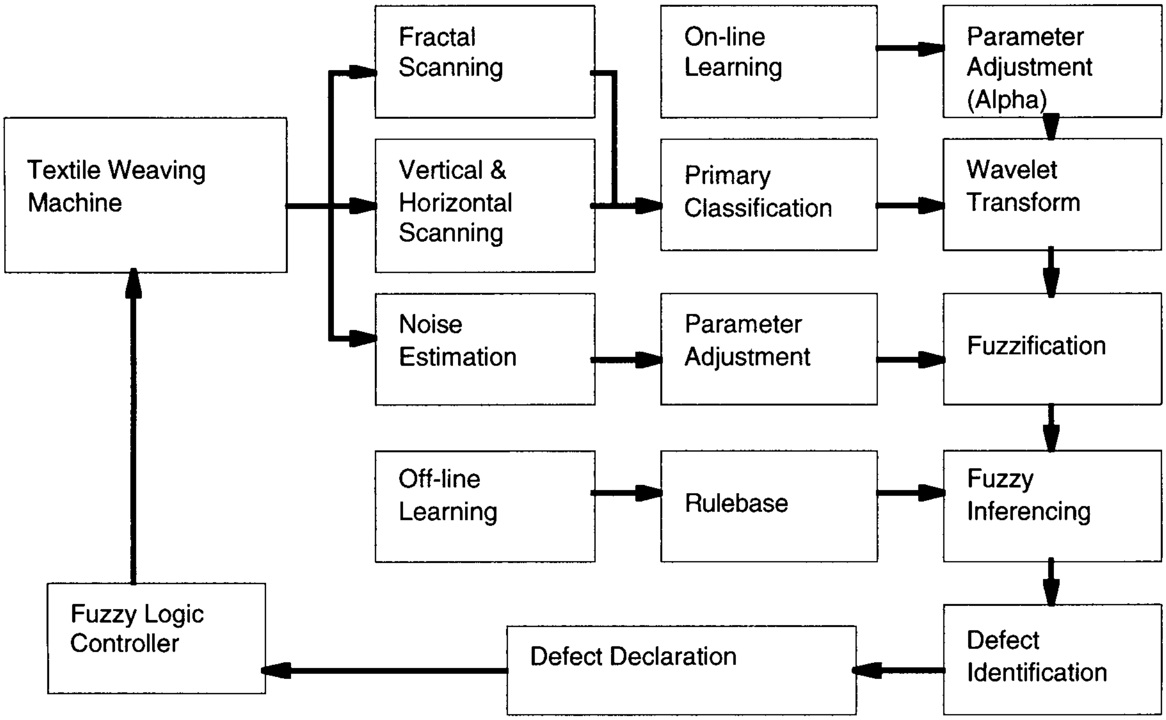

Textile Weaving Machine
Fractal Scanning
Vertical & Horizontal Scanning
Noise Estimation
Off-line Learning
On-line Learning
Primary Classification
Parameter Adjustment
Rulebase
Parameter Adjustment (Alpha)
Wavelet Transform
Fuzzification
Fuzzy Inferencing
Defect Identification
Defect Declaration
Fuzzy Logic Controller

Figure 1. SYSTEM ARCHITECHTURE

module that trains the defect identification mechanism and an on-line system that employs the trained mechanism to carry out real-time identification tasks. Figure 1 depicts the overall hardware/software architecture. A number of CCD array cameras are used to capture fabric images and transfer them to a CPU through appropriate interfacing devices. The images are processed in the CPU (typically a PC or a DSP processor) where features are extracted and defects identified. Upon defect identification, the defect type is displayed and relevant information is passed on to the operator while control signals may be generated and used to control the loom so that certain consistent defects may be reduced or eliminated.Pre-processing routines are usually designed to prepare the image for further analysis. Among them, filtering, image enhancement and segmentation may be exploited to enhance the contrast between the target image and background clutter or noise. Of special interest to the fabric inspection process is a series of projection techniques that reduce the two-dimensional image into a stream of one-dimensional data. We employ horizontal, vertical and fractal projections that respond efficiently to warp, filling and area defects, respectively. These scanning tools not only reduce considerably the computational burden, but also assist in the class identification process since each scanning method is maximally sensitive to one specific category of defects. After scanning, feature extraction and classification algorithms are called upon to complete the inspection procedure.

3.2 Architecture of an Intelligent Defect Identification Algorithm

The defect detection and identification problem is in fact a problem of classifying features into different categories. It may be viewed as a mapping from the feature space to the decision space. Several techniques have been proposed in the literature for feature extraction and defect of fault classification. In the "intelligent" arena, fuzzy logic classification techniques based on the fuzzy e-means algorithm, multi-level architectures for feature classification exploiting fuzzy logic again and fuzzy rule bases have been proposed as candidate technologies. Most of the intelligent techniques employ a learning mechanism (unsupervised or supervised) which uses information from an expert, historical data, extrinsic conditions, etc., to train the model. Neural net based detection methods use learning to adjust the weights of individual neurons while fuzzy associative memories employ learning to design the inferencing hypercube.

We are introducing in this paper a novel methodology based upon a combination of fuzzy logic and wavelet transforms. The Fuzzy Wavelet Analysis (FWA) technique attempts to accommodate some of the shortcomings of the existing methods of feature extraction, defection detection and identification. The proposed method allows one to extract features from various signals and to optimize them for use in the knowledge base. This requires both on-line and off-line learning. Finally, in order to demonstrate the validity of the technique, an application example relating to textile fabric inspections is given below.

3.3 Feature Extraction

Features or signatures are characteristic signals representing a particular operating mode of a system. Feature extraction is accomplished by processing the input data so that the presence of the operating modes is accentuated. An effective feature extraction methodology ensures a more robust identification process. However, feature extraction has been the most difficult part of any identification scheme. This is due to the fact that normal operational modes and noise have signatures that are very similar to feature signatures. The FWA employs the wavelet transform (WT), with different wavelet functions, to extract features from the signal x(j). This is illustrated in Figure 2. The wavelet transform generates wavelet coefficients which are employed by a fuzzy inference engine; to seek a match between the current state of the wavelet coefficients and the templates stored in the knowledge base. A wavelet is represented by $y_{a,b}$, where y is the mother wavelet, a is a scaling factor and b is the translation in time or space depending upon the context. $n \in N$ number of wavelet scales are chosen which produce best results for the anticipated defects. The set of wavelets for some b is given by .

$$D=\left\{\Psi_{a_1 b},\Psi_{a_2 b},\Psi_{a_3 b},\cdots,\Psi_{a_n b}\right\}$$

The wavelet scales a_i, i = 1,...,n, are obtained via an optimization process that works off-line. The wavelet coefficients, for each wavelet ya,b and the input signal x(j), are calculated as

$$c'_{a,b} = \sum_{j=0}^{N} x(j)\psi_{a,b}(j)$$

where N is the number of samples for which ya,b is non-zero. Since most of the features produce a signature in a wide range of frequencies that is spread over a range of time (or space), m > 0 number of coefficients are buffered. A transformation x is applied to the wavelet coefficients to obtain a trend of the input signature

$$c_{ji} = x(c'_{a_i b_j}) \qquad i=1,\ldots,n \quad j=1,\ldots,m$$

The transformation x varies from one application to another. It may envelope extraction, magnitude of a complex value, etc. The transformation x involves scaling so that

$$\max\{c_{ji}\} \le 1.$$

The coefficients c_{ij} are stacked in a matrix arrangement which is referred to as the Information Matrix W, since it stores the information about all the features under examination.

3.4 Defect Classification

Fuzzy Wavelet Analysis combines fuzzy tools and wavelet transform techniques for providing a robust feature extraction and failure detection and identification scheme. The input signal first undergoes preprocessing and then features are extracted using the wavelet transform. The extracted features are fuzzified and an inference engine used the knowledge-base to declare fault conditions. The fuzzification process adapts dynamically to external disturbances so that the classification performance is continuously improved. The knowledge base stores the information employed in the decision making process. It attempts to organize all the information available about the system (the relationship between defect features and defect modes) through heuristics, input-output data or any other information source. The knowledge-base of the FWA contains the representative features for different wavelet scales, for each one of the anticipated defects, and its development is carried out off-line. An optimization procedure chooses the best set of wavelet scales (ai) and stores the template features for these scales.

The fuzzy rule-set is given by the following statements:

$$\textit{if } w_1 \textit{ is } F_1^1 \textit{ and } w_2 \textit{ is } F_2^1 \textit{ and.....}$$
$$\textit{......and } w_n \textit{ is } F_n^1 \textit{ then } y \textit{ is } G^1$$
$$\textit{if } w_1 \textit{ is } F1^2 \textit{ and } w_2 \textit{ is } F_2^2 \textit{ and .....}$$
$$\textit{......and } w_n \textit{ is } F_n^2 \textit{ then } y \textit{ is } G^2$$

An important attribute of intelligent systems is their ability to restructure or update the knowledge-base, based on information from past experience. This is termed as learning. The FWA uses both on-line and off-line learning to increase the knowledge about the system and improve the detection and identification processes. On-line learning attempts to nullify the effects of variation in lighting, imperfection of the CCD sensor element or spherical aberration of the lenses. It accomplishes this by adjusting the FWA parameters either during pre-processing or at the time of fuzzification. The off-line learning process provides the FWA with the ability to generate its knowledge-base from sample defects. This is an optimization process which selects the best values for the given set of defects and stores the corresponding features in the rulebase. The defect detection and identification algorithms are implemented on a PC when process speed is low or on special DSP processors when speed is of concern.

4. TESTING

The first stage of prototype testing was done in the weaving laboratory at Georgia Tech on an airjet weaving machine with excellent results. The second stage of testing was carried out at a textile plant weaving industrial fabric on a projectile loom also with most encouraging results. As a result, a licensee of the technology was found and the commercialization began. A much smaller and less expensive unit resulted and this also was tested at Georgia Tech.

Results of tests done on polyester/cotton sheeting fabric are shown in Table 1. This fabric has a plain weave construction of 40 ends and 40 picks per centimeter. This is indeed a difficult fabric for a vision system and defects appear quite prominent in some areas and almost invisible in others. This is especially true in defects such as the misdraw, misreed and mispick.

The first level of detection is determined by whether the peak value of signal-to-noise in the convolved signals exceed a given threshold. The second level of detection involves summary logic which considers the position of the defect and "tracks" detection at the first level to account for some frames which may not indeed indicate a defect in that section of fabric. This is to mimic the human grader's taking a global view and realizing that a defect may "come and go", where the human grader recognizes it as the same defect which is continuing.

In this way the summary logic can account for detection rates as low as 33% on the first level of detection and still correctly call a continuing defect. In like manner, an isolated defect that barely exceeds the threshold will generally be ignored as an isolated anomaly. Thus the last column indicating nearly perfect detection.

Figure 2. Fuzzy Wavelet Analysis

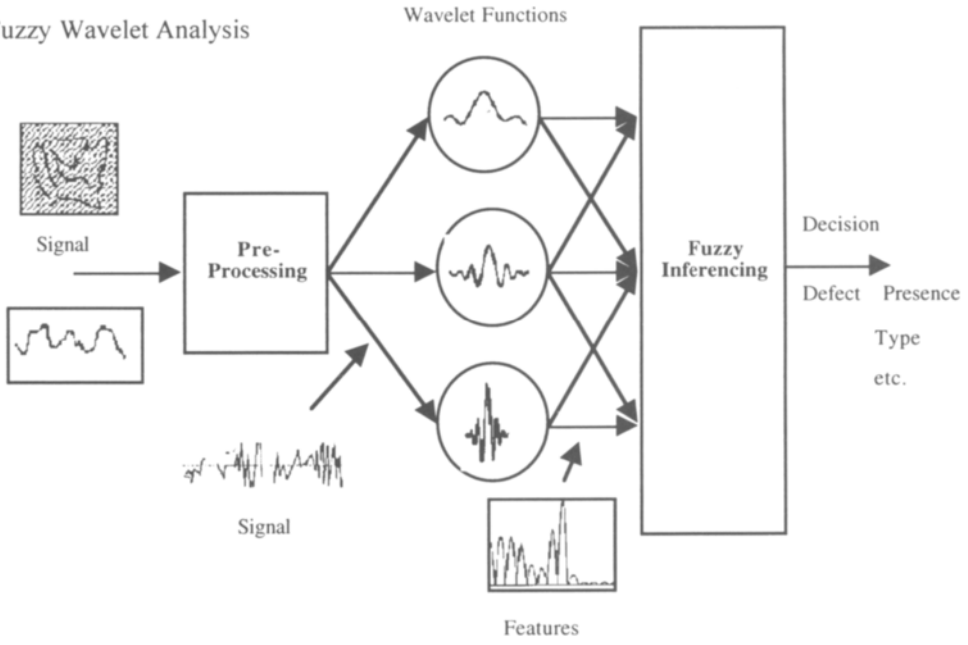


Table 1. Results of testing of sheeting fabric

The columns in the table have the following meanings:

Defects- Name of defect
% of Frames- Percent of camera frames correctly called at the first level of detection.
# Correct- Number of camera frames correctly called at the first level
Total- Total number of frames containing the defect
% Detection- Percentage of defects called correctly after the second level of detection.

	Defects	% of Frames	# Correct	Total	% Detection **
Warp	Endout	93	200	215	99+
	Misreed	43	40	94	99+
	Slack End	100	1	1	99+
	Misdraw	33	4	12	99+
	Double End	91	178	196	99+
Weft	Mispick	54	7	13	99+
	Thin Place	100	5	5	99+
	Thick place	100	8	8	99+
	Start Mark	100	8	8	99+
	Slubby Filling	80	4	5	99+
	Broken Pick	100	3	3	99+
Area	Oil Spot	100	2	2	99+
	Grease Spot	100	1	1	99+
	Woven-in-Waste	100	2	2	99+
No Defect		89	569	636	99+

** Detection Rate using summary logic

5. CONCLUSIONS

The increasing availability of low cost imaging and digital processing hardware has made the dream of automatic fabric inspection a reality. The great interest shown and the investments already made by the leading textile companies indicate that the need is evident.

REFERENCES

1. D. C. Mead, H. Kasdan and J. L. Dorrity, "Method for Automatic Fabric Inspection", U. S. Patent No. 4,124,300 ,November 1978.
2. H. Sari-Sarraf, “Vision System for On-Loom Fabric Inspection”, IEEE/IAS Textile, Fiber & Film Conference, May 1998.
3. J. L. Dorrity and G. Vachtsevanos, “On-line Defect Detection for Weaving Systems” IEEE Textile Fiber & Film Conference, Atlanta, Ga., May1996.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Estimation of Workpiece Quality by Using State Observer

M. Shiraishi and S. Aoshima
Faculty of Engineering, Ibaraki University
12-1 Nakanarusawa 4 chome, Hitachi, 316 Japan

Multiple state of workpiece condition in turning such as dimensional error and surface roughness are estimated by using a state observer and compared with the results of measuring instruments. For this aim, the machining process is simplified and modified to derive the dynamical model for these estimates. Experimental results show that the change of instantaneous depth of cut approximately indicates the dimensional error and the estimates of cutting force agree well with the results by a tool dynamometer.

1. Introduction

There has been growing interest in recent years in the possible up-grading of in-process monitoring systems in manufacturing through the integration of multiple sensors. This increasing trend towards automation and high precision, as well as the increasing costs of raw materials and labor, has created a great need in industry for a comprehensive approach to the problem of product quality assurance. The final goal of in-process measurement is to provide information on the basis of which corrections can be made during machining to ensure that the workpiece is manufactured to the desired size. The success of in-process measurement strictly depends on the sensor itself. A number of in-process sensors have been developed so far and, unfortunately, none of them are practically applicable to the real production floor.

With this background into consideration, this paper describes a way to estimate the instantaneous depth of cut and cutting force using a state observer [1]. These factors are directly related to surface roughness and/or dimensional error. Thus, sensorless in-process measurement of machine qualities is possible. Results obtained using a UMM (Universal Measuring Microscope) and a tool dynamometer are compared with the estimates and the effectiveness of the proposed method is confirmed through several experiments.

2. Modeling of Machining Process

The structure of a single-point turning operation with regenerative chatter is modeled by a linear dynamic system with instantaneous and time-delayed feedback, as shown in Fig. 1. The cutting process dynamics is expressed by the cutting stiffness K, and the workpiece-machine tool structure is expressed by the second-order transfer function as depicted in the figure. Km is the workpiece-machine tool structure stiffness, ξ is the damping ratio , ω_n is the natural frequency of the workpiece-machine tool structure, F is the cutting force, and Uo is the stationary depth of cut. Here, U is the instantaneous depth of cut, μ is the overlap factor, and T is the time required for one revolution of the workpiece. Cutting force plays an important role in workpiece quality such as surface roughness, and this force is affected by the instantaneous depth of cut, which changes with the variation of cutting process parameters. The changes of U and F are thus treated as the disturbances in the estimation process under consideration.

Assume that $Udis$ and $Fdis$ are, respectively, the disturbances of instantaneous depth of cut and cutting force that may be related to dimensional error and/or surface roughness. To estimate these disturbances separately, the machining process shown in Fig.1 is simply modified into two processes with only a feedforward path as indicated

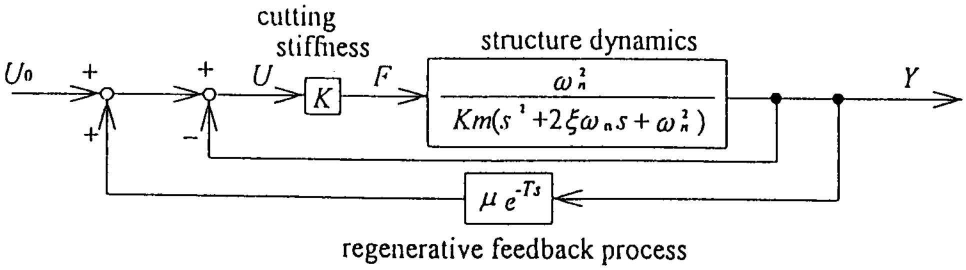
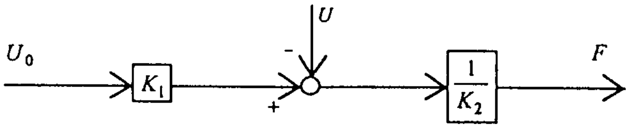
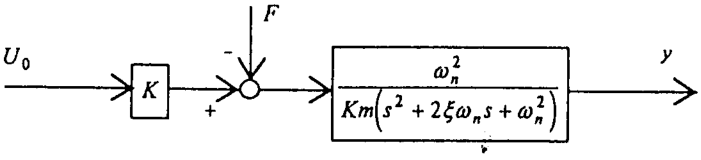

cutting stiffness
structure dynamics
Uo
+
+
+
−
U
K
F
$\frac{\omega_n^2}{Km(s^2+2\xi\omega_n s+\omega_n^2)}$
Y
μe^{-Ts}
regenerative feedback process

Fig. 1 Structure of machining process


U_0
K_1
U
−
+
$\frac{1}{K_2}$
F

(a) model for the estimation of Udis


U_0
K
F
−
+
$\frac{\omega_n^2}{Km(s^2+2\xi\omega_n s+\omega_n^2)}$
y

(b) model for the estimation of Fdis

Fig. 2 Simplified model for the estimation

in Fig. 2 [2]. In Fig. 2(a), U is assumed to relate to the change in width of the cut, which is called an outer modulation. $K1$ is a parameter denoting the relationship between the stationary and instantaneous depths of cut. In other words, the effects of the regenerative path in Fig. 1 are included in the coefficient $K1$ which changes within 1.0.

Figure 2(b) shows a simplified model from input Uo to output y with the disturbance of F. From Fig. 2(a), the dynamics of depth of cut is expressed as

$$K2F = K1Uo - U \tag{1}$$

where $K2$ is the compliance of the cutting process ($K2 = 1/K$). To derive the disturbance $Udis$, let us consider the actual values $K1$ and $K2$ in Eq. (1) with the nominal values $K1n$ and $K2n$:

$$K1 = K1n + \Delta K1 \tag{2}$$

$$K2 = K2n + \Delta K2 \tag{3}$$

where $\Delta K1$ and $\Delta K2$ denote the errors between the actual and nominal values. Substitution of Eqs. (2) and (3) into Eq.(1) yields

$$K2nF = K1nUo - Udis \tag{4}$$

$$Udis = U - Uo\Delta K1 + F\Delta K2 \tag{5}$$

Equation (4) is the model for the instantaneous depth of cut during machining and $Udis$ denotes the disturbance of depth of cut estimated by the state observer. If the errors $\Delta K1$ and $\Delta K2$ in Eq. (5) are negligibly small, $Udis$ is expressed by U. This indicates that if the parameter errors in the machining process are sufficiently small, the disturbance $Udis$ is nearly equal to the instantaneous depth of cut.

In the same manner, the dynamics of cutting force is derived from Fig. 2(b) as

$$\frac{Km}{\omega_n^2}\ddot{y} + \frac{2\xi Km}{\omega_n}\dot{y} + Kmy = KUo - F \tag{6}$$

and this can be rewritten

$$K3\ddot{y}+\varsigma\dot{y}+K4y=KUo-F \qquad (7)$$

where $K3=Km/\omega_n^2$, $\varsigma=\varsigma n+\Delta\varsigma$, and $K4=Km$.

For the parameter change of

$$K3=K3n+\Delta K3,\quad \varsigma=\varsigma n+\Delta\varsigma$$

$$K4=K4n+\Delta K4,\quad K=Kn+\Delta K$$

the model of the cutting force disturbance is given by

$$K3n\ddot{y}+\varsigma n\dot{y}+K4ny=KnUo-Fdis \qquad (8)$$

$$Fdis=F+\Delta K3\ddot{y}+\Delta\varsigma\dot{y}+\Delta K4y-\Delta KUo \qquad (9)$$

where $K3n, \varsigma n, K4n$ and Kn are nominal values and $\Delta K3$, $\Delta\zeta$, $\Delta K4$, and ΔK denote the parameter errors between actual and nominal values. Equation (8) yields $Fdis \cong F$ when these errors are negligibly small. The disturbance $Fdis$ is thus estimated by a state observer also.

3. Design of the Observer

3.1. Estimate $\hat{U}dis$ of U

$Udis$ in Eq. (4) is simply estimated through the low-pass filter, as shown in Fig. 3. The parameter $T1$ must be determined by taking the frequency bandwidth of estimation into account.

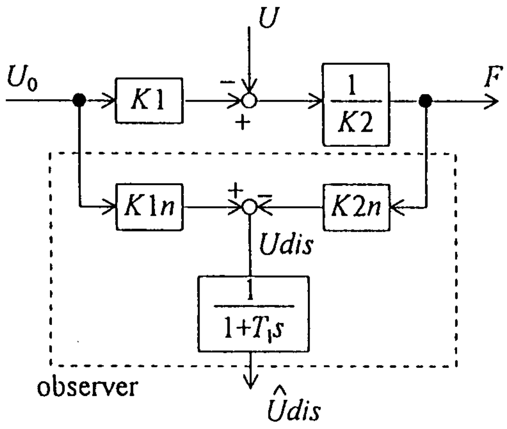


Fig.3 Obsever for $\hat{U}dis$

3.2 Estimate $\hat{F}dis$ of F

From Eq. (8), we obtain the following equation with the state variables of $y=x_1$, $\dot{y}=x_2$, and $Fdis$, providing that the disturbance $Fdis$ is kept constant during the sampling period.

$$\begin{bmatrix}\dot{x}1\\ \dot{x}2\\ Fdis\end{bmatrix}=\begin{bmatrix}0 & 1 & 0\\ -\frac{K4n}{K3n} & -\frac{\varsigma n}{K3n} & -\frac{1}{K3n}\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}x1\\ x2\\ Fdis\end{bmatrix}+\begin{bmatrix}0\\ \frac{Kn}{K3n}\\ 0\end{bmatrix}Uo$$

$$y=[1\,0\,0][x_1\ x_2\ Fdis]^T \qquad (10)$$

Since the system (10) is observable and the output y can be measured by the displacement sensor, a dynamical model of the second-order observer is designed as

$$\dot{z}=\begin{bmatrix}-\frac{\varsigma n}{} & -\frac{1}{K3n}\\ -L2 & 0\end{bmatrix}z+\begin{bmatrix}-\left(\frac{\varsigma n}{K3n}+L1\right)-\frac{L2+K4n}{K3n}\\ -L1L2\end{bmatrix}y+\begin{bmatrix}\frac{Kn}{K3n}\\ 0\end{bmatrix}Uo \qquad (11)$$

$$\hat{x}=\begin{bmatrix}0 & 0\\ 1 & 0\\ 0 & 1\end{bmatrix}z+\begin{bmatrix}1\\ L1\\ L2\end{bmatrix}y$$

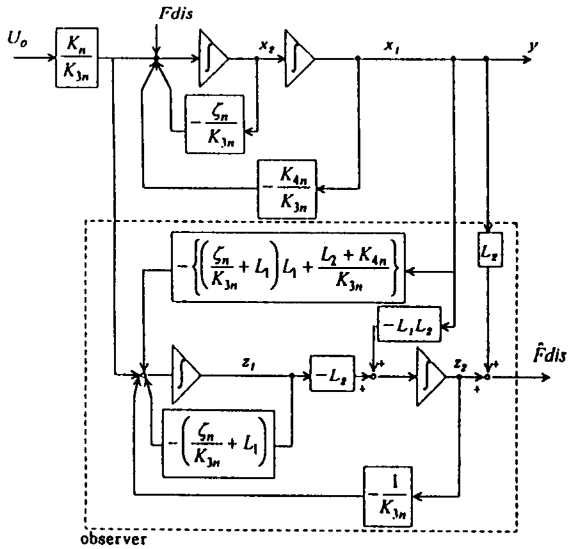


Fig.4 Observer for $\hat{F}dis$

where L_1 and L_2 are the design parameters. Figure 4 represents the overall state variable diagram for the estimate $\hat{F}dis$.

3.3 Actual System Design

To determine a dynamic characteristic of the cutting process, several experiments were performed by the input of a step function, which is given by a sudden increase of depth of cut. The response curves obtained were analyzed with a personal computer. Because of the nature of the machining system, it is not easy to accurately determine machine parameters such as ω_n, K, Km, ξ, and μ in Fig. 1. These parameters change with cutting conditions and are functions of various movable elements of the machine tool. The representative identified values are f_n=41.5 Hz (f_n: frequency), $K1$=0.8, ξ=0.3, and K / Km = 0.5 at a revolution time T=0.1 s (i.e., spindle speed = 550 rpm).

As for Fig. 3, $K1n$=0.8 and $K2n$=16.8 μ/kg. Thus,

$$\hat{U}dis = \frac{1}{1+0.1s}\left(K1n - K2nF\right) \tag{12}$$

As for Fig. 4, assigned poles of the observer are -6 and -8 in the analog system design. This leads to

$$\dot{z} = \begin{bmatrix} -1.7\times10^4 & -17.1\times10^5 \\ -24 & 0 \end{bmatrix} z + \begin{bmatrix} -41.2\times10^5 \\ -192.1 \end{bmatrix} y + \begin{bmatrix} 13.6\times10^4 \\ 0 \end{bmatrix} U0, \quad \hat{x} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} z + \begin{bmatrix} 1 \\ 8 \\ 24 \end{bmatrix} y \tag{13}$$

Thus

$$\hat{F}dis = z_2 + 24y$$

$$\dot{z}_1 = -1.7\times10^4 z_1 - 17.1\times10^4 z_2 - 41.2\times10^5 y$$

$$\dot{z}_2 = -24z_1 - 192y + 13.6\times10^4 Uo$$

4. Experiment and Results

The effectiveness of the proposed method was confirmed in experiments under the cutting conditions of cutting speed 100-200 m/min, feed rate 0.1-0.25 mm/rev, depth of cut 0.2-0.6 mm, and cutting tool P30. The experimental setup is shown schematically in Fig. 5. The experiments were conducted on a bench-type lathe using S45C carbon steel workpieces (40 mmϕ × 450 mm) supported by the tail stock. A relative motion, which is denoted by $x_1(=y)$, between the workpiece and the tool was measured by an eddy-current-type transducer attached to the tool post. A tool dynamometer was also used to measure the cutting force F. In the preliminary experiments, machine vibrations including chatter were found to be between 30 to 180 Hz for the range of cutting conditions. Therefore filtering was achieved with the band-pass of 20 to 200 Hz to reduce the noise level. The

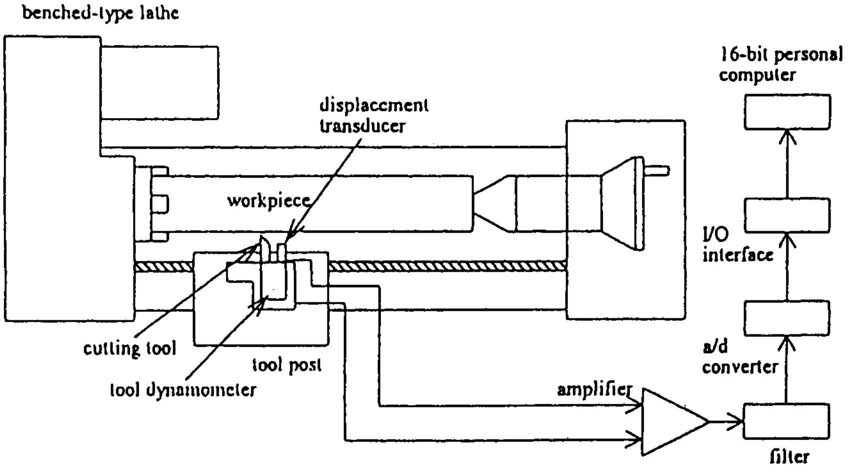


Fig. 5 Experimental setup

signals filtered through a/d converters were fed to the 16-bit personal computer to calculate the parameters and the observer coefficients.

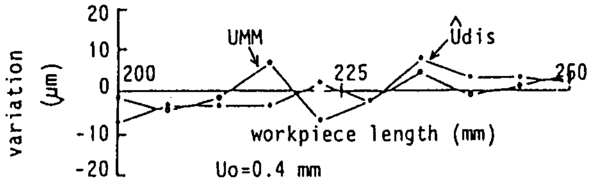
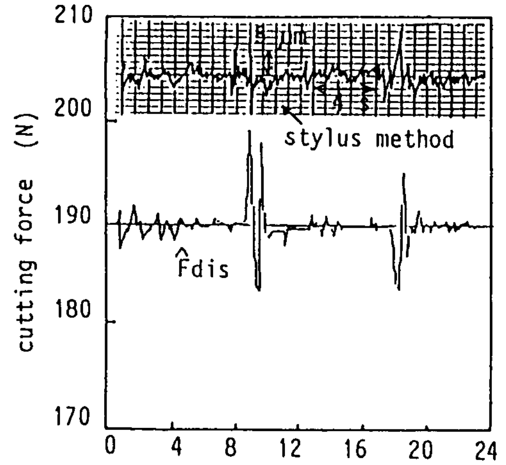
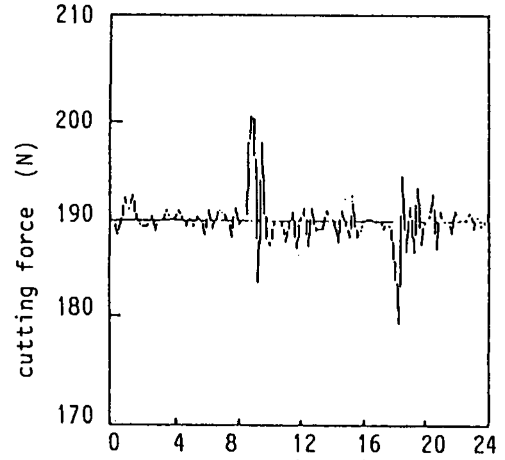


Fig.6 Results of estimation $\hat{U}dis$ and measurement by UMM



(a) estimation $\hat{F}dis$



(b) measurement by the tool dynamometer

Fig.7 Estimation and measurement of cutting force

The instantaneous depth of cut estimated by the observer $\hat{U}dis$ when the stationary depth of cut Uo is 0.4 mm is shown in Fig. 6. Unfortunately, there is no tool to measure U during cutting. In the figure, the measurement with a UMM (Universal Measuring Microscope) after machining is also depicted and is compared with the result of $\hat{U}dis$ at a length of 200 mm from the start of the cut. Although both measurements indicate slightly different traces at some local points, the estimated $\hat{U}dis$ is approximately the same trend as that measured by the UMM. This shows the feasibility of, for example, using $\hat{U}dis$ to estimate the dimensional error.

In Fig. 7, the estimate of the cutting force $\hat{F}dis$ is shown together with the measurements made using the tool dynamometer. Surface roughness measurement with the stylus method is also demonstrated in Fig. 7 where the trace was obtained at the same workpiece position, 300 mm from the start of the cut. The change of the stylus measurement corresponds approximately to $\hat{F}dis$.

5. Conclusion

Instantaneous depth of cut and cutting force were estimated using state observers separately. To do this, the machining process with a regenerative path was simplified and modified into two feedforward processes. In particular, the instantaneous depth of cut, which cannot generally be measured using the sensing technique, approximately indicated the trend of dimensional errors. The effectiveness of the proposed method was thus confirmed by the experiments using a tool dynamometer and the stylus instrument.

References

1. S. Kawaji, and Y. Suenaga, Control of Cutting Torque in the Drilling Process Using Disturbance Observer, 1995ACC, WM8, pp.723-728, 1995.
2. M. Shiraishi, and H. Fujita, H., Optimal Control of Chatter in Turning, Int. J Machine Tools Manuf, Vol. 31, No. 1,pp.155-167, 1991.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

ON THE POSITION AND VELOCITY MEASUREMENTS OF THE HIGH SPEED PISTON OF A FREE PISTON ENGINE

Erkki Lehto, Tapio Virvalo and Juha Inberg

IHA, Tampere University of Technology
P.O.Box 589, 33101 Tampere, Finland, e-mail elehto@cc.tut.fi, virvalo@cc.tut.fi, i139347@cc.tut.fi

Abstract: Last fifteen years many attempts have been made to get hydraulic free piston engines to work also in practise. The main idea has been to put together directly mechanically, without a crankshaft, a hydraulic piston and the piston of a Diesel engine. In principle there are different structures and combinations of hydraulic pistons and Diesel engine pistons used in hydraulic free piston engines. In this paper basic structures of hydraulic free piston engines are presented shortly. Some of them benefit and drawbacks are also discussed. In this paper we concentrate so-called double piston hydraulic free piston engine. In a Diesel engine (engine with a crankshaft) the injection timing is based on the mechanical movement of a crankshaft. In a hydraulic free piston engine the right injection moment has to be based on the measurement or evaluation of the piston movement. The capability to measure both position and velocity of a piston is one of most important things in the development work of a free piston engine. The circumstances where these measurements should be done are very heavy. In this paper some potential possibilities to realize both position and velocity measurement systems are presented. Their benefits and drawbacks are discussed.

Keywords: Free Piston Engine, Position measurement, Velocity measurement

1. INTRODUCTION

The quite old idea of a free piston engine (FPE) has got now and then attention in different research institutes. The basic idea was presented at 1673 by a Dutch Christian Huygens. In the 18th century many attempts were made to use the free piston engine principle with air and gas. At the beginning of the 19th century hot gas generators based on the free piston engine idea were used almost all over the wold. The basic idea was applied also on air compressors quite largely in this century. During last thirty years some attempts have been made to develop electrical generators based on the free piston engine idea. One of the first patents on a hydraulic free piston engine (HFPE) was presented at early 1960's. Later versions remind quite a lot that one. Last twenty years many attempts have been made to get hydraulic free piston engines to work also in practise. The main idea has been to put together directly mechanically, without a crankshaft, a hydraulic piston and the piston of a Diesel engine or an Otto engine. The Diesel engine is the most common basic engine used in research and development works of HFPEs. In principle there are different structures and combinations of hydraulic pistons and Diesel engine pistons. The following basic structures have been under the R&D work: a single piston engine [1], an opposed piston engine [2], [3] and a dual piston engine [4].

In a Diesel engine (engine with a crankshaft) the injection timing is based on the mechanical movement of a crankshaft or the rather simple electrical measurement of its movement. In a HFPE the right injection moment has to be based on the measurement of the piston movement. In principle the same situation as in crankshaft engine might be achieved in a HFPE, if the position of the piston could be measured or evaluated somehow near the top dead centre (TDC)

2. SINGLE PISTON ENGINE

The principal structure of a single piston HFPE is depicted in Figure 1. This construction principle is the simplest of the HFPE alternatives because there are only one combustion piston and one hydraulic piston.

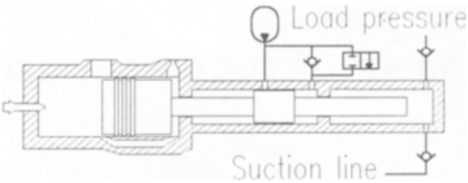


Figure 1 The principal structure of a single piston hydraulic free piston engine.

In addition a special compression piston and an accumulator are used to achieve a compression stroke of a Diesel engine piston. The main advantages of this type of HFPE are the control easiness of its working frequency and a relative easy way to start the engine. The starting and timing of a compression phase can be controlled by the control valve between compression piston and accumulator. This means also that the actual volume flow of the hydraulic pump can be controlled in a quite large range, when an accumulator is used in the suction and high pressure sides.

The main drawbacks of this type are the special system required to achieve the compression stroke and the dynamical unbalance in the working condition.

3. OPPOSED PISTON ENGINE

The principal structure of an opposed piston HFPE is depicted in Figure 2. In this engine there are two combustion pistons but only one combustion chamber. Hydraulic part consists of two different parts. There are two pumping pistons and two compression pistons.

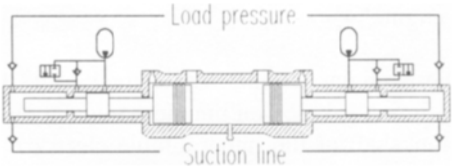


Figure 2 The principal structure of an opposed piston hydraulic free piston engine.

The main advantage of an opposed piston HFPE is that the engine is dynamically balanced since opposed moving masses compensate each other. Because there are separated compression pistons, a little bit same kind as in a single piston HFPE, the controlling of a working frequency might be possible. The main serious problem is to synchronize the movements of two separated pistons. In practice this requires some special valves and careful timing. The required compression system complicates also this type of a HFPE.

4. DUAL PISTON ENGINE

The principal structure of a dual piston hydraulic free piston engine is depicted in Figure 3. There are two separated combustion chambers at both ends of an engine. The combustion pistons are connected to each other by a rod in which hydraulic pistons are integrated in the middle. A remarkable characteristic of this HFPE type is that there is no compression piston and the system required to achieve the compression work. The compression stroke in the other combustion chamber is performed with the help of the combustion energy of the other combustion chamber.

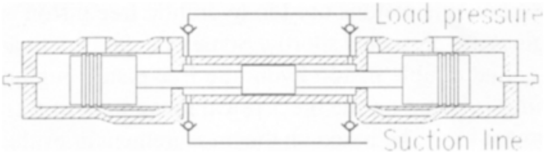


Figure 3 The principal structure of a dual piston hydraulic free piston engine

The main advantage of a dual piston HFPE is that no separated compression system is needed. Its power density might be highest of these three types of HFPE. The main drawbacks of this type are that it is not dynamically balanced and that there are only limited possibilities to control the working frequency.

5. CHARACTERISTICS OF A DUAL PISTON ENGINE

In this paper we concentrate a dual piston HFPE construction. In this engine there are two piston units put together as is shown in Figure 3. The hydraulic pump works in both movement directions and when one Diesel piston is in a combustion phase the other one is in a compression phase. Because two Diesel pistons are working together without any really mechanical movement limiting links, the control of a whole engine is very critical. In every stroke some energy balance has to be achieved otherwise the engine does not work properly. The ways to influence the energy balance are to control the injection timing and fuel quantity and the suction and delivery side valves of a hydraulic pump. Some main characteristics of all HFPEs are that the combustion process is very fast and the velocity and acceleration of the engine piston are very high. Without the mechanical movement limiting devices, the control of an engine stroke is based on right timing of injection and hydraulic valves.

5.1 Piston velocity and acceleration.

The following basic calculation of the velocity and acceleration of a dual piston HFPE can be made based on typically dimensions and characteristics of HFPEs under development in different R&D groups. The power level of the Diesel engine part is about 25 kW. The diameters of combustion pistons are in the range of 90...110 mm. The stroke lengths vary in the range of 90... 170 mm. The maximum working frequencies are 30...40 Hz. The symmetric movement profile can be assumed for a dual piston HFPE, because in both direction there are same kinds of compression and expansion phases, Figure 4.

The following specifications are used: Working frequency f= 40 Hz, Stroke Lp=105 mm, Total cycle time Tc=25 ms, Dead time T1=0.2 ms and Stroke time T2=12.4 ms

Assuming the trapezoid movement profile so that 80% of the stroke time is used for constant velocity and 10% for both acceleration and deceleration the maximum velocity v_{max} is got Eq.(1).

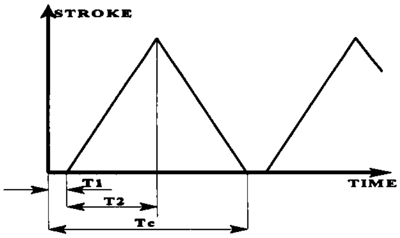


Figure 4 Piston stroke as a function of time

$$L_p=0.5*v_{max}*T_{acc}+v_{max}*(T2-T_{acc}-T_{dec})+0.5*v_{max}*T_{dec}$$

$$v_{max}=\frac{Lp}{T2-0.5*0.2*T2}=\frac{0.105}{0.012-0.0012}=9.7\ m/s \tag{1}$$

where L_p is the stroke length of the engine [m], T_{acc} is the acceleration time [s], T2 is the stroke time [s], T_{dec} is the deceleration time [s].

The calculated maximum velocity corresponds quite well with a prototype measured velocity. It is however significantly lower than about 20 m/s reported in /1/ for a single piston HFPE. The corresponding maximum velocity of the piston of a crankshaft engine is about 13 m/s.

The acceleration and deceleration of the piston a_{max} with assumed movement profile and time distribution is calculated with Eq.(2)

$$a_{max}=\frac{v_{max}}{0.1*T2}=\frac{9.7}{0.0012}=8100\ m/s^2 \tag{2}$$

The corresponding acceleration/deceleration of the piston of a crankshaft engine is 3300 m/s^2 and in a single piston HFPE it is more than 20000 m/s^2 /1/.

6. CONTROL PRINCIPLE OF HFPE

Two main problems in a dual piston HFPE from the control point of view are that an engine becomes "unstable" or it stops to run. If the kinetic energy of the moving of the whole piston structure becomes too high there is a danger that combustion pistons impact on the cylinder heads of the combustion chambers. If the kinetic energy is too low, the pressure in combustion chamber remains too low and the burning of the fuel cannot happen.

The basic force equation of the piston is presented in Eq.(3)

$$m_{pis}*\ddot{y}_{pis}-A_{Lcopi}*p_{Lcopi}+A_{Rcopi}*p_{Rcopi}-A_{pump}*p_{hyd}-F_{fric}=0 \tag{3}$$

where m_{pis} is the mass of the whole piston structure [kg], y_{pis} is the position of the piston [m], A_{Lcopi} is the area of the left-hand side combustion piston [m^2], p_{Lcopi} is the pressure of the left-hand side combustion chamber [MPa], A_{Rcopi} is the area of the right-hand side combustion piston [m^2], p_{Rcopi} is the pressure of the right-hand side combustion chamber [MPa], A_{pump} is the area of the hydraulic pump piston [m^2], p_{hyd} is the load pressure [MPa] and F_{fric} is the friction force [N].

According to Eq.(3) there are many parameters which influence the kinetic energy of the whole piston structure. A control system can adjust some of these parameters. There are at least the following ways to control and influence the kinetic energy of the piston structure: a) adjusting an amount of fuel injected in both combustion chambers, b) controlling the timing of the fuel injection, c) varying the pressure level in the working side of a hydraulic pump, d) varying the pressure level in the suction side of a hydraulic pump and e) using active valves in working and suction sides of a hydraulic pump and timing their operation.

If some control activities have to take place, the main problem is how to realize it in the correct moment and on the correct level. In traditional crankshaft engines it is very easy to synchronize for instance the injection to the position and a combustion piston. The compression ratio is independent on the load in a crankshaft engine, but in a HFPE it depends quite strongly on the loading condition. If good efficiency and low pollution level are important, the compression ratio should be kept in quite narrow limits.

One way to try to solve these problems is to measure the position and velocity of the piston. The requirements to the measurements and control of these parameters are significantly higher in dual piston HFPEs than in single piston and opposed piston HFPEs. In single piston and opposed piston HFPEs there are some possibilities to adjust the stroke length of a combustion piston at the bottom dead centre (BDC). Then almost constant compression phases can be guarenteed.. In dual piston HFPEs there are not this kind of possibility.

7. REQUIRED MEASUREMENTS

In a Diesel engine (engine with a crankshaft) the injection timing is based on the mechanical movement of a crankshaft or a rather simple electrical measurement its movement. In a HFPE the right injection moments have to be based on the measurements of the piston movement. In principle the same situation as in crankshaft engine might be achieved in a HFPE, if the position and velocity of the piston could be measured near the top dead centre (TDC). An alternative way to achieve right timing and fuel injection is to use some measurements and model-based evaluation of piston velocity and position.

The capability to measure both the position and velocity of the piston is one of the most important things in the development phase of a HFPE. The circumstances in which these measurements should be done are very heavy. The velocity (8...20 m/s) and acceleration (7000...20000 m/s^2) of a piston are very high, the temperature range is large and there are very high vibrations in a HFPE. Typically the acceleration of an engine body might be about 600...1000 m/s^2 with amplitude about 15...25 mm. This problem occurs in a single piston and dual piston HFPE, but not in an opposed piston HFPE.

At the end of a compression phase the compression pressure increases very rapidly as a function of the piston position so the requirement to know the position of the piston is relatively high. Due to high velocity the follow-up capability of a position measurement system should also be high. It can be evaluated that the accuracy of position measurement should be 0.1 mm and the repeatability 0.05 mm. The accuracy requirement of the velocity measurement system is not so critical, but the repeatability should be reasonable and the response should be quite high.

Because there might be many control activities and monitoring requirements the controller of a HFPE should be realized in the digital way.

8. PROPOSED VELOCITY AND POSITION MEASUREMENT METHODS

In this paper some possibilities to realize both position and velocity measurement systems are considered. The TDC of the piston position has to be measured quite accurately so that the actual compression ratio of the combustion process could be detected and the right injection moment determined. One way to fulfill these requirements is to use several magneto-resistive sensors, Figure 5.

Using two sensors in opposite sides of the piston small variations of the radial position of the piston (in y direction) can be compensated. The electronic

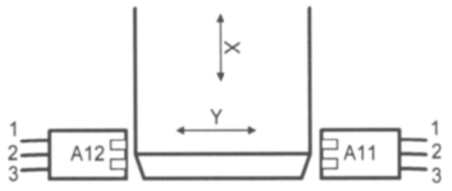


Figure 5 Two magneto resistive sensors (type Siemens) near by skirt of a Diesel piston.

circuit of the sensors produces an output voltage proportional to the piston position (in x direction), Figure 5. The developed electronics of the sensors produces an output voltage proportional to the piston position, Figure 6. The ferromagnetic piston skirt influences the magnetic field strength of the sensors. Resistance values of both sensor elements 1-2 increases when the piston skirt approach the sensors. The changes in the field strength influence sensitive resistors which in turn influence the balance of the measurement bridge. Unfortunately magneto resistive sensors have a great negative temperature coefficient that has an significant effect on the output voltage of sensors, Figure 7.

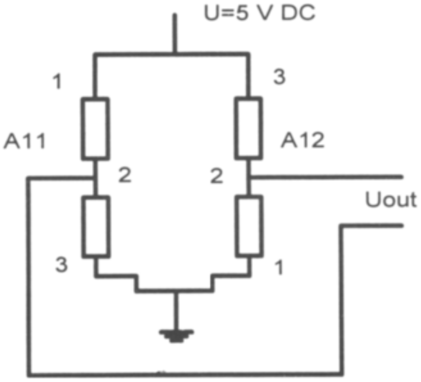


Figure 6 Wheatstone bridge circuit of the magneto-resistive elements.

Usable measurement range depends on how piston skirt has be bevelled. When bevelling is 5 mm long and 2 mm deep practicable measurement movement is about 8 mm. Measurement accuracy 0.1 mm is within reach when normal backlash of the piston is less than 0.3 mm and compensation have being well done electrically and programmatically One another measurement principle could be made by many magneto resistive or hall-effect sensors one after another in long direction. Their comparative changes on the characteristic values are entering position value..

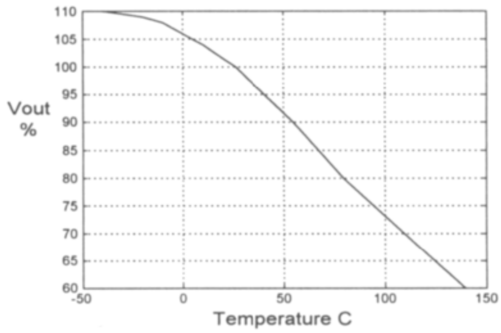


Figure 7 Output voltage of a magneto resistive element versus temperature

To achieve better timing of the injection and to evaluate the control requirements of the next injected fuel amount the velocity of the piston is also measured with a couple of variable reluctance magnetic pick-up coils, Figure 8. The pick-up sensors are placed in the middle range of the piston movement. The distance of the sensors is known accurately and DSP-based control system is able to measure time with very good resolution and also accurately, Figure 9. The distance between pick-up sensors is 20 mm. Because the velocity of a dual piston HFPE varies in the range -10...10 m/s it is important how accurately the outputs of sensors can be detected. The output signal of a pick-up sensor

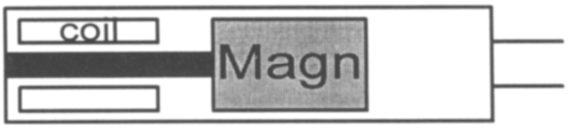


Figure 8 Variable reluctance magnetic pick-up

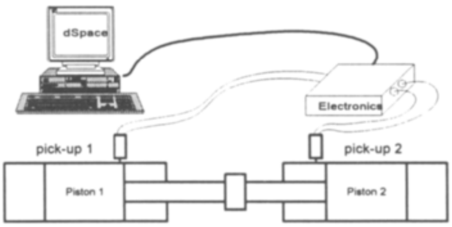


Figure 9 The placement of pick-up sensors

varies according to the velocity of the piston, Figure 10. If a threshold voltage is set at 2.5 V the variation of a detected piston position is about 3 mm. With the arrangement of sensors depicted in Figure 9 the actual distance between sensors varies significantly less. More important is the repeatability of sensors and their individual differences.

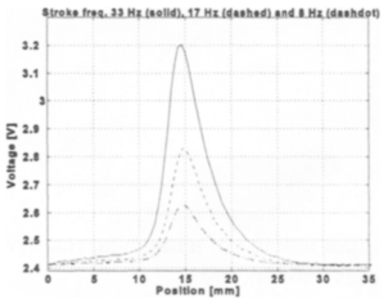


Figure 10 Variation of a pick-up sensor output as a function of piston velocity

The accuracy of the time measurement has a great influence on the operation of HFPE. Sampling period of a control system should be short enough. The conversion time of a a/d-converter sets limit for a sampling time. The velocity of the piston could also be measured with hall-sensors. The output of a hall-sensor is TTL-compatible and it could be measured using digital input, which is faster.

9. SUMMARY

Hydraulic free piston engines are very promising alternative to some special applications. In the development phase of a HFPE it is quite necessary to make many kind of measurement to find out suitable value ranges of control parameters. Environment conditions are very hard. Temperature varies quite large range and vibration of an engine body is very strong. Some potential sensor candidates have been found for both position and velocity measurements. Initial test of these sensors and measurement systems are going on and main efforts will be put to find sensors that could tolerate the hard condition of a HFPE.

REFERENCES

1. Somhorst J. And Achten P., The Combustion Process in A DI Diesel Hydraulic Free Piston Engine. SAE paper 960032, pp. 37-50.
2. Hibi A. and Hu Y., A Prime Mover of A Free Piston Internal Combustion Hydraulic Power Generator and a Hydraulic Motor. SAE paper 930313, 10 p
3. Beachley N. and Fronczak F., Design of a Free Piston Engine-pump. SAE paper 921740, 10 p.
4. Sampo M., Power Aggregate. European patent, Patent nr. 0280000 (1988)

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Inference Mechanisms Applied to On-Line Electro-Pneumatic Fault Detection and Diagnosis

D. S. Evans

Department of Engineering, University of Wales College, Newport,
P.O.Box 180, Newport, South Wales, UK, NP9 5XR

Tel: +44(0)1633 432451
Fax: +44(0)1633 432430
E-mail: d.evans@newport.ac.uk

This paper reflects on previous developments and limitations in PLC based electro-pneumatic fault detection utilising linear actuator timing strategies. A proposal for an on-line dynamic pressure prediction model is subsequently presented utilising a modular based neural network architecture. The application of a fuzzy-neural inference mechanism is also considered as the basis for a component specific diagnostic strategy.

1. INTRODUCTION

The rapidly changing industrial climate in the UK actively subscribes to the notion that each facet of any commercial manufacturing operation be closely examined with a view to obtaining the best product quality combined with efficiency optimization. Manufacturing profitability and efficiency are clearly closely linked and in this context continuous plant condition monitoring can play a vital role. In many such operations, PLC and computer based on/off sequential electro-pneumatic systems make a substantial contribution to the production environment since pneumatic drives offer many attractive engineering features such as high power to weight ratio, high range of operating speed, insensitivity to overload and relatively low cost. These systems can, however, be highly interactive and even a single component failure can often induce more than one symptom making the fault finding process an arduous and sometimes dangerous task. These failures inevitably occur at the most inconvenient time imposing technical, organisational and financial difficulties. Hence, on-line supervision and monitoring of such systems with particular emphasis on predicting the future occurence of faults in addition to the detection of sudden faults must be seen as a highly desirable and achievable goal. Plant maintenance programmes would subsequently be enhanced with substantial improvements in the safety function combined with increased availability of the industrial process for the production function. Surveys in industrial and process automation indicate that around 80% of electro-pneumatic failures are due to sensors, valves, actuators or associated wiring with the remaining 20% of failures being attributable to I/O modules, CPU's and power supplies [1].

Fault diagnosis is often required to be component specific or at the very least sub-system specific. This normally requires sensors to be placed within the appropriate circuit and often not in the ideal place due to operating constraints, particularly near the actuators. The majority of electro-pneumatic systems, however, have little or no provision for fault detection and diagnosis since arguments based on through-life costs in support of on-line monitoring equipment rarely gain support when set against additional capital expenditure.

A number of approaches to fault detection exist such as performing checks on measurable process signals, using high-low alarm limits and quality control charts. However, the information provided by these methods is limited in relation to the cause of the fault and the detection of faults due to degradation [2]. A large class of fault detection techniques use observers and Kalman filters to estimate state variables [3]. An alternative to this method is parameter estimation [4] which utilises a process model, the parameters of which are repeatedly estimated and monitored for deviation from their normal boundary limits.

These techniques all aim to generate either a signal, a model parameter, or a process coefficient which is then monitored for time variation within statistical limits. Deviation outside these limits constitutes the detection of a fault. The mechanics of progressing from fault detection to fault diagnosis, however, requires a good knowledge of the system under consideration and is the process of determining the cause of a problem based on observable symptoms.

2. PREVIOUS WORK

At UWCN, the actuator performance during pre-defined cyclic operations has been the centre of investigation since any fault condition in the system will always instigate recognisable actuator symptoms. Typical of these symptoms are:

- Incorrect speed
- Incorrect thrust
- Incorrect sequence
- Actuator arrest

Whichever fault or combination of faults within the system has produced the symptom, the fundamental problem must always be associated with either flow, pressure or direction of actuator travel. Changes in any of these parameters affect the timing of each actuator as it moves or attempts to move between its two extreme limit positions. Hence, work has focused on the development of timing strategies for the detection of these adverse symptoms [5,6] based on the actuator timing diagram shown in Figure 1. The technique utilises the existing proximity detectors found at the extreme positions of the actuator stroke. These are primarily used to provide feedback for synchronisation of the actuator control sequence. However, they also enable the PLC to monitor the progress of the actuators during the extension or retraction phases from which appropriate timing data can be obtained.

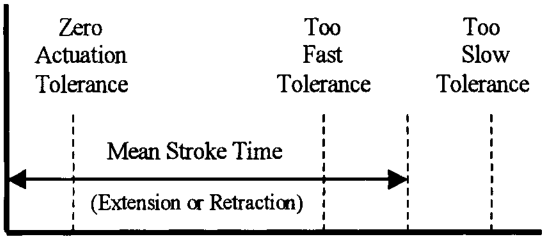


Figure 1. Actuator Timing Diagram.

This data is analysed in conjunction with the logic states of the proximity detectors in order to identify possible deviations from normal actuator timing. The following information can subsequently be established and presented by the PLC:

- Actuator identification
- Actuator symptom
- Actuator direction

Results have been encouraging giving consistently reliable results using various PLC monitoring programmes. However, the technique does not extend to the identification of specific faults once an adverse symptom has been identified.

When a fault occurs in an electro-pneumatic system operating in a production environment, it is normal practice to replace the faulty element in its entirety with a "sound" element in order to reduce down-time and enhance reliability. Although most system elements consist of a series of sub-systems such as solenoid actuated directional control valves, the important consideration must always be faulty element identification, removal and replacement, with subsequent re-commission and production start-up. Failures which can be experienced by electro-pneumatic systems fall into two main categories:

- Sudden failure associated typically with valve seizure, actuator seizure or sensor failure resulting in system arrest

- Progressive failure associated typically with component degradation or air leaks

The sudden failure mode can be detected using the actuator timing strategy previously outlined. Once the arrested actuator has been identified by the PLC together with its current state in the production cycle, the manual troubleshooting exercise to identify the fault is relatively straightforward. However, it is the progressive mechanical failures due to component degradation which can be difficult to detect and which introduce inefficiencies into the operating performance of the system. The mechanical fault producing progressive system failure cannot be identified directly by utilising the timing strategy alone. However, one of the main demands of this research initiative is to utilise minimal instrumentation in order to reduce initial capital outlay. This problem is currently being addressed by considering Artificial Neural Networks to aid the troubleshooting process. ANN's are being increasingly used as decision support tools in a wide range of applications and they have the potential for modelling the complex behaviour of highly non-linear pneumatic elements. Hence, work is currently being directed towards combining the actuator timing strategy with a neural network strategy in order to aid the identification of specific faulty mechanical or distribution elements contributing to progressive system failure.

3. CURRENT WORK

3.1. Neural Networks

Neural networks are now being exploited in situations which cannot justify detailed mathematical modelling and associated validation [7,8,9] including fluid power systems [10,11,12]. One advantage is that network training can be undertaken using real component data such as pressures and volumetric flow rates inclusive of data for the appropriate fault condition. Multilayer Perceptrons (MLP's) are the most flexible and commonly used signal processing tool, with successful pattern matching applications in many diverse areas of work including fluid power [13]. Neural networks can offer advantages for this application since they do not require an in depth knowledge of the systems' dynamic behaviour. Typically, neural networks have been employed for the detection of leakage in hydraulic systems [14].

Fault finding in electro-pneumatic systems normally involves the testing of pressure at strategic points in the system. However, this requires suitable pressure tapping points and associated instrumentation which this research seeks to avoid. The approach, therefore, is to train a neural network to form a non-linear, dynamic model of the electro-pneumatic system capable of predicting system pressure. The input parameters to the network which are currently being assessed are:

- Actuator Stroke-Time
- System Supply Pressure
- External Actuator Load
- Actuator Direction

The load, supply pressure and stroke-time data for both extension and retraction actuator phases is currently being amassed and will be used to train a multilayer perceptron neural network using supervised learning based on error back-propagation. Validation and assessment procedures on how well the network generalises with previously unseen data will subsequently be initiated.

A typical electro-pneumatic model configuration indicating selected points for pressure prediction is shown in Figure 2 and the corresponding mulilayer perceptron with appropriate inputs and outputs is shown in Figure 3:

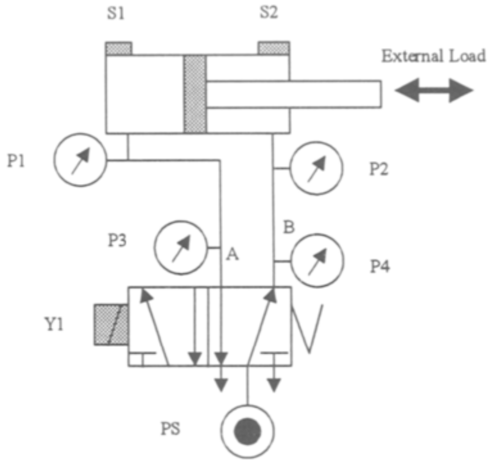


Figure 2. Electro-Pneumatic Model.

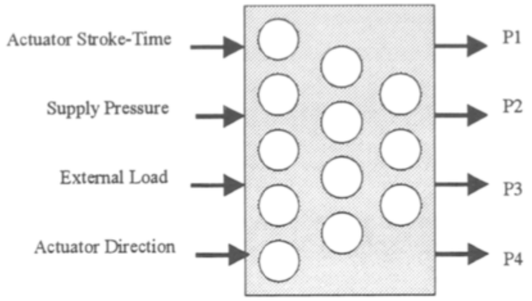


Figure 3. Multilayer Perceptron.

3.2. Modular Neural Network Strategy

The development of a single neural network to accommodate the operating characteristics of an integrated electro-pneumatic system introduces serious problems due to their size and complex architecture. The network will have been suitably trained to match outputs with inputs for an existing set of components in a system array. Each type of pneumatic component, however, such as the actuator and directional control valve, has its own dynamic performance characteristics. Even 'identical' components will exhibit different performance due to variations in, for example, machined tolerances, stiction and sub-assembly torque settings. Subsequently the introduction of a replacement component into the system will necessitate on-line network retraining with inherent loss of production. In the case of pneumatic systems, which themselves exhibit highly integrated modular structures, it would be useful if the corresponding architecture of neural networks incorporated for fault finding also consisted of a multitude of integrated networks. The various knowledge areas associated with individual pneumatic components would be distributed in multiple coupled modular neural network architectures. Each module network would be designed to model a single pneumatic element and could be individually trained off-line using customised learning algorithms. In addition, the type of module network could be selected to cope most easily with the type of pattern transformation required such that each component would have its own individual network with customised architecture. The block diagram architecture of a typical modular network structure is shown in Figure 4 where the outputs of the individual expert networks are mediated by a gating network. Each module relates specifically to the individual pneumatic and distribution elements incorporated in the electro-pneumatic model shown in Figure 2.

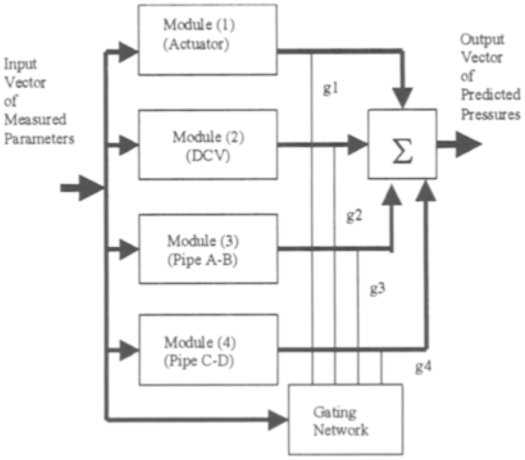


Figure 4. Modular Neural Network.

A modular network offers several advantages over single neural networks in terms of the following parameters:

3.2.1. Speed of Learning

If a complex function is naturally decomposable into a set of simpler functions, then a modular network is able to learn the set of simpler functions faster than a multilayer perceptron is able to learn the undecomposed complex function.

3.2.2. Data Representation

The representation of input data for a modular network tends to be easier to understand than the corresponding data for a multilayer perceptron.

3.2.3. Hardware Constraints

With a limit is imposed on the number of neurons used in an artificial neural network, the representation of multi-dimensional spaces may be distributed amongst multiple networks, each able to compute different functions.

3.3. Test Programme

The facility currently being employed for the acquisition of neural network training and validation data is shown in Figure 5:

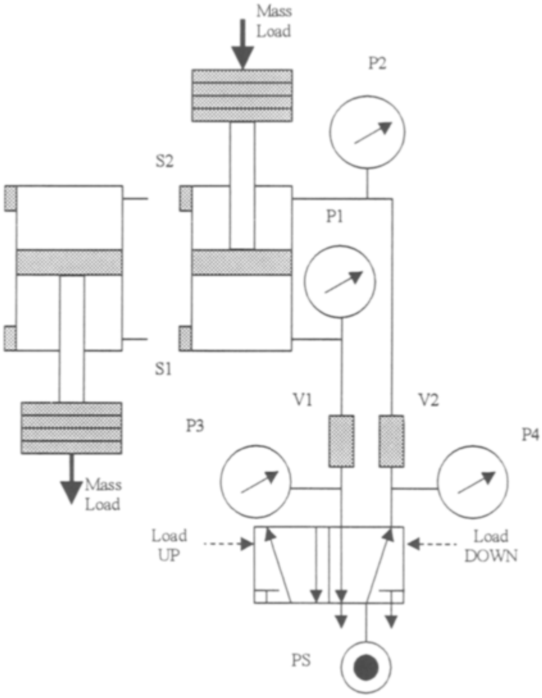

Mass Load
P2
S2
P1
S1
V1
V2
P3
P4
Mass Load
Load UP
Load DOWN
PS

Figure 5. Pneumatic Test Facility.

The actuators are arranged in a vertical plane to allow for the provision of a simple but accurate mass load arrangement for both actuator extension and retraction. Proximity sensors S1 and S2 provide start/stop signals for an electronic timer in order to determine actuator stroke times under varying load, back-pressure and supply pressure conditions. Metered-out speed control is incorporated using valves V1 and V2 and a 5/2 double pilot directional control valve is used for actuator extension and retraction. The system is provided with a stable pressure source, PS, and a constant ambient temperature is arranged for each test procedure.

4. PROPOSED FURTHER WORK

Neural Networks are not able to perform logic-like rules because the distribution of connection weights in the network are almost impossible to be interpreted in terms of IF-THEN rules. Various approaches have been proposed to incorporate fuzzy logic elements into neural networks [15,16]. The resulting network topolgy is then able to perform fuzzy inference rules through analysing the values of the connection weights. Hence it is proposed to incorporate a fuzzy rule base in the pressure prediction neural network to generate a fuzzy neural network (FNN) diagnostic inference engine based on the model configuration shown in Figure 6:

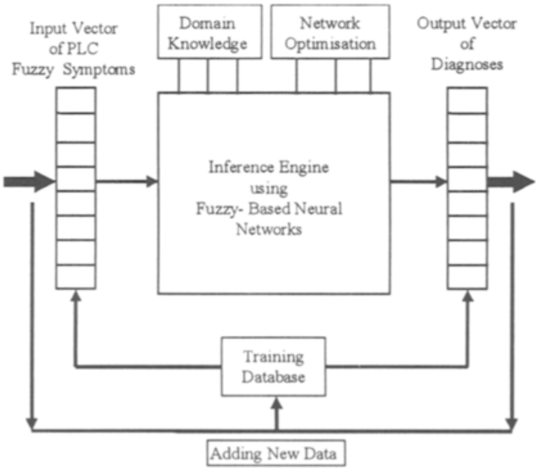

Input Vector of PLC Fuzzy Symptoms
Domain Knowledge
Network Optimisation
Output Vector of Diagnoses
Inference Engine using Fuzzy- Based Neural Networks
Training Database
Adding New Data

Figure 6. Fuzzy-Neural Diagnostic Model.

The proposed model consists of a centralised PLC providing the monitoring and control function for an on-off sequential electro-pneumatic hardware system. At the front end of the diagnostic facility, an interface will receive various symptoms from the PLC, each symptom consisting of an attribute and its associated linguistic properties. Initial attributes will be:

- Actuator Stroke -Time
- Actuator Load
- Supply Pressure

Typical linguistic properties for the Actuator Stroke-Time attribute will be:

- Negative Large
- Negative Medium
- Negative Small
- Positive Small
- Positive Medium
- Positive Large

Based on its acquired knowledge, it is planned that the inference engine will suggest a single or set of fault conditions with different confidence levels. At the output interface, these fault conditions will be interpreted to provide an ordered list of the most likely diagnoses.

5. CONCLUSIONS

The application of a timing strategy in PLC based electro-pneumatic systems for the purpose of identifying adverse actuator symptoms has produced consistently reliable results. The development of a multilayer perceptron and modular based neural network architectures to predict on-line system air pressures appears promising but work is still proceeding. These predicted pressures will form the basis of a component specific diagnostic strategy where it is proposed to extend the work by adding a fuzzy rule base to the neural networks. This development will be specifically directed towards the identification of individual pneumatic components undergoing progressive failure due to degradation.

REFERENCES

1. Norris D., “Smart Distributed Systems: Distributed Control for Factory Floor Automation”, FieldComms'95, Conference Proceedings, pp.5/1-5/14, Nov.7-8, Hinckley, Leicester, UK,1995.
2. Urwin P., “Fault Detection and Diagnosis in Industrial Systems”, Joint Hungarian-British International Mechatronics Conference, September 21-23, 1994, Budapest.
3. Watanabe K., Himmelblau D.M., “Incipient Fault diagnosis of Non-Linear Processes with Multiple Causes of Faults”,Chem. Eng. Sci., Vol 139, No.3, 1984.
4. Isermann R., ”Process Fault Detection Based on Modelling and Estimation Methods-A survey”, Automatica, 1984, Vol 20, No.4.
5. Evans D.S. and Underwood P., "Self-Diagnostics in Pneumatic Systems using Programmable Logic Controllers", Joint Hungarian-British International Mechatronics Conference, September 21-23, 1994, Budapest.
6. Evans D.S., "Condition Monitoring of Industrial Fluid Power Systems using Programmable Controllers", International Conference on Recent Advances in Mechatronics, ICRAM'95, August 14-16, 1995, Istanbul, Turkey.
7. Hunt, K. J., Sbarbaro, O., Zbikowski, R, and Gawthrop, P. J., Neural Networks for Control Systems-a survey, Automatica, 1992, 28(6), 1083-1112.
8. An, P. E., Brown, M., Harris, C. J. and Chen, S., S., Comparitive aspects of neural networks for on-line modelling of dynamic processes. Proc. Instn. Mech. Engrs., Part 1, 1993, 207(14), 223-241.
9. Pham, D. T., Neural networks in engineering, First National Conference on Applications of AI techniques in engineering, Naples, October 1994, pp. 3-36.
10. Sanada K., Kitagawa A. and Pingdong Wu., "An application of a neural network to adaptive control of a servo-system". In Fluid Power, Proceedings of the 2nd JHPS International Symposium on Fluid Power, (Edited by Maeda T.), pp. 303-308. E. &F.N. Spoon, London (1993).
11. Newton D. A., "Design and implementation of a neural network controlled electro-hydraulic drive". Proc.IMechE 208. 31-42, (1994).
12. Xu, P. at al, Feasibility of training a neural network based hydraulic component simulator using the complex method. Seventh Bath International Fluid Power Workshop, Bath, September, 1994.
13. Pham, D. T. Neural Networks in engineering. In Application of Artificial Intelligence in Engineering (Eds G. Rzevski, R. A. Adey and D. W. Russell) Vol. IX, 1994 (Computational Mechanics Publications).
14. Stewart J.C., The application of artificial intelligence to fault detection in hydraulic cylinder drive systems, PhD thesis, University of Wales Cardiff, 1995.
15. Pedrycz W. and Rocha A.F., "Fuzzy-set based model of neurons and knowledge-based networks," IEEE Trans. Fuzzy Syst., vol 1, pp. 254-266, Nov., 1993.
16. Sanchez E., "Fuzzy logic knowledge systems and artificial neural networks in medicine and biology" in An Introduction to Fuzzy Logic Applications in Intelligent Systems. Norwell MA: Kluwer, 1989, ch.11.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

An Unsupervised Artificial Neural Network Approach to Adaptive Noise Cancellation Applied to On-line Tool Condition Monitoring

Mark Girolami[a, & b] and James Findlay[c]

[a] Laboratory for Open Information Systems, Brain Science Institute, RIKEN, 2-1 Wako-Shi, Japan.

[b] Department of Computing and Information Systems, University of Paisley, High Street, PA2 1BE.

[c] Department of Mechanical Engineering, University of Paisley, High Street, PA2 1BE.

Abstract

Enhancing detected machining signals such as cutting force is an important research field in automatic manufacturing systems. For a milling machine, the problem is particularly difficult because of multiple sources caused by the numerous cutting teeth. The problem is further compounded in that the multiple sources are distorted and combined at the sensors in a non-stationary and unknown manner. Multivariable adaptive noise cancellation (MVANC) techniques have been employed to reduce the desired signal distortion; however, certain restrictions such as noise alone periods and their detection is necessary. This paper reports on the practical application of an unsupervised artificial neural network (ANN) model to this particular noise reduction and signal enhancement problem. This method allows the noise reduction to proceed in a 'blind' and unsupervised manner.

1. INTRODUCTION

On-line mechanical system monitoring and machine tool condition diagnoses are the key factors towards the development of the unmanned automatic machining system. The diagnostic correction of the condition monitoring system is degraded when the measured signal is corrupted by the surrounding noise.

Several noise canceling techniques can be applied to improve the signal-to-noise ratio (SNR) of a diagnostic signal, e.g. power spectrum subtraction and coherent filtering. These techniques are limited because they either rely on a synchronizing signal or are effective in reducing stationary noise only. A method, which is applicable to stationary or non-stationary signals, is adaptive noise cancellation (ANC) [1]. This particular technique does not require *a priori* knowledge of the diagnostic signal. It does however require the following set-up conditions to be valid 1) The primary input must contain both the desired signal and the background noise. 2) The noise at the primary and reference inputs must be correlated and none of the desired signal is allowed at the reference input. These constraints require that there is no cross-channel leakage between the primary and reference sensors. For the particular problem under investigation this condition is unlikely to be satisfied. Introducing noise alone periods during which the transfer functions of the interfering signals can be estimated [2] can circumvent the problem of cross-channel leakage. This however introduces the additional requirement of having knowledge of when cutting and non-cutting periods occur.

A method of enhancing the diagnostic signal in the presence of non-stationary interfering noise is desired such that the signal separation will continue irrespective of whether the machine is in cutting mode or not. This necessitates a blind filtering technique where the sensor output will always be an unknown mixture of the desired and unwanted signals. Recent research advances in the field of unsupervised learning have seen the development of biologically inspired filters which make this form of blind signal separation a possibility [3].

The paper is structured as follows; Section 2 presents the artificial neural network model for blind signal separation along with the maximum likelihood adaptation learning rule. Section 3 describes the experimental conditions of the trials reported here. Section 4 reports the results of the noise reduction experiments employing the blind method for noise suppression. Section 5 concludes the paper and provides a discussion on these preliminary results and considers further potential work.

2. BLIND SIGNAL PROCESSING

The accepted model for an observed signal at, for example, the output of a sensor is given as

$$x_i(t)=\sum_{j=1}^{M}\sum_{k=0}^{\infty}a_{ij}(k)s_j(t-k) \tag{1}$$

The source signals, which are assumed to be generated from independent processes, are given as $\mathrm{s}(t)=[s_1(t)\dots s_M(t)]$. The transmitting medium is modeled as a linear causal filter of infinite length $\{a_{ij}(k)\}_{k=1}^{\infty}$ and this has a distorting effect on the signal denoted by the convolution. It is assumed that the sensor is strictly linear and so the observed signal is a linear summation of each convoluted independent source.

The problem of recovering the original sources, or preferentially suppressing the unwanted noise terms, is essentially one of inverse filtering. There is no *a priori* knowledge of the transmitting filters or the source signals. The filtered output of the network will provide estimates of the underlying sources based on the chosen inverse filter model. This can be denoted, for all *i*, as

$$\hat{s}_i(t)=\sum_{j=1}^{M}w_{ij}*x_j(t) \tag{2}$$

Where * denotes convolution. A finite length anti-causal filter can generically approximate the required inverse filter. The technique detailed in this paper is based on strictly causal finite impulse response (FIR) filters and so will provide a stable inverse for strictly minimum phase filter models. The recurrent ANN model employed for the blind signal separation is given in Figure 1.

2.1 Network Adaptation For Separation

From Figure 1, the output of each node at time *t* for an $M \times M$ network with memory based synaptic weights of length *L* is given as

$$y_i(t)=x_i(t)+\sum_{j\neq i}\sum_{k=1}^{L}w_{ij}(t-k)y_j(t-k) \tag{3}$$

Where the subscripts denote spatial relations between the nodes within the network and the (*t-k*) terms denote delays of *k* samples from time *t*. It is detailed in [4] that the network can perform blind signal separation by adapting based on maximum likelihood estimation (MLE) of the parameters of the source signal distributions.

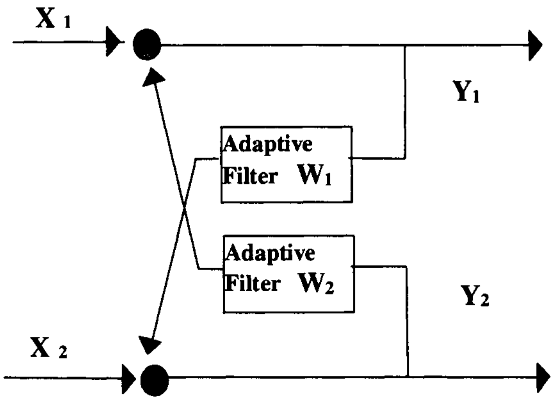


Figure 1. A Dual-Channel Recurrent ANN Signal Separation Model.

The generic MLE is given as

$$\hat{\theta}_{ML}=\arg\max\left(N^{-1}\sum_{i=1}^{N}\log p(\mathbf{x}_i;\hat{\theta})\right) \tag{4}$$

Whose value is given by the solution of

$$\frac{\partial}{\partial\hat{\theta}}\left(N^{-1}\sum_{\mathrm{i}=1}^{\mathrm{N}}\log p(\mathbf{x}_i;\hat{\theta})\right)=0$$

This corresponds to a network adaptation algorithm [4] of

$$\Delta w_{ij}^{n+1}(t-k)=\mu_n f\left(y_i^{n+1};w_{ij}^{n}(t-k)\right)y_j(t-k) \tag{5}$$

and the sequential update follows

$$w_{ij}^{n+1}(t-k) = w_{ij}^{n}(t-k) + \Delta w_{ij}^{n+1}(t-k) : \forall k$$

Where the nonlinear term appearing in (5) is the score function of the output distribution parameterized as the underlying source signals.

$$f(y_i) = p_i'(y_i)/p_i(y_i) \tag{6}$$

The noise characteristics can be described by a Normal distribution; it is a standard result that the corresponding score function is simply a precision-scaled linear function of the dependant variable. So for the case where we make the valid assumption of Gaussian distributed noise then

$$f(y_i) = -y_i/\sigma_i^2$$

This then provides a decorrelation criterion for (5) that is

$$\Delta w_{ij}^{n+1}(t-k) = -\mu_n y_i y_j(t-k)/\sigma_i^2 \tag{7}$$

Inspection shows that the cutting signal has a multi-modal distribution. A simple model, which can be employed, is proposed in [5] and provides a score function of the form

$$f(y_i) = -y_i + \tanh(y_i)$$

Inserting this nonlinear term into the weight update equation then gives a nonlinear decorrelation criterion which takes the non-Gaussian nature of this signal into account.

$$\Delta w_{ij}^{n+1}(t-k) = -\mu_n(y_i - \tanh(y_i))y_j(t-k) \tag{8}$$

The use of (7) in one channel will preferentially suppress the noise term and likewise (8) will suppress the cutting signal within the channel. There is no requirement for a noise alone period to estimate the noise transfer function. The 'blind' nature of this adaptive learning allows the separation of the mixed signals that have overlapping spectra [4].

This form of blind adaptation has been successfully applied to the enhancement of speech in the presence of noise [4]. As the method is blind, in that no *a priori* knowledge of the signals or the coupling mechanisms is required, this provides a less constrained ANC method for on-line condition monitoring.

3. EXPERIMENTAL CONDITIONS

The experimental conditions reported here are similar to the two-channel case reported in [2]. The experimental tests were performed on a Bridgeport Series 1 industrial milling machine. The set-up conditions are detailed in Table 1.

Table 1. Experimental machine set-up

Bridgeport Series 1	Specification
Cutting Tool	V cutter, 90° inc. 3" x½"
Feed rate	15.24 mm/min
Spindle revolution	150 rpm
Cutting depth	1mm
Applied workpiece	High carbon steel (BS970: EN30B)

The experimental work in this paper has employed vibration sensors to obtain the desired cutting signal and the interfering noise. The vibration signals were measured using two accelerometers (Brüel and Kjær type 4371). In [2] several configurations of sensor montages were investigated. Accelerometers were mounted on both the motor and gear housings, and also the workpiece holder. In this instance the desired signal is the cutting signal, as a result the accelerometer positions on the motor and gear housings were not optimum, as they did not provide a clear cutting signal. Positions on the workpiece holder can provide a clear cutting signal; however, there are several difficulties in adopting these positions:

1. It is not always possible to mount the accelerometers on the workpiece holder.

2. The workpiece holder on the milling machine is continually in motion and as a consequence the probe connections may be disturbed during the machine operation.

Therefore in practice, these accelerometer mounting positions may not be ideal. However, the positions on the motor and gear housings are useful when investigating motor bearing or gear faults. Since the cutting signal is the desired signal, to obtain the clearest cutting signal the accelerometers are mounted around the cutting spindle.

The noise corrupted cutting signals are sampled from the accelerometers at 20kHz. A charge amplifier is required for each accelerometer to pre-amplify the detected signal that is generated by piezoelectric discs inside the accelerometer.

3.1 Adaptive Noise Cancellation

When the ANC approach is taken to the reduction of the non-stationary corrupting noise the following procedure is followed. First, noise alone periods are introduced to the noise canceller for noise transfer paths identification. Secondly, when the error output has been minimized, the desired cutting signals along with the corrupting noise are presented to the noise canceller.

This adaptive approach has been found to provide signal to noise improvements of 4.69dB on average [2] taken over a number of cutting cycles. This approach, whilst adaptive has one drawback when considered for this application. The noise alone periods required to identify the noise transfer paths impact the efficiency of the machining process in that the tool requires to be removed from the cutting surface. Although this can be accommodated in a part program it is not desirable for reasons of cutting efficiency.

A multiple channel approach [2] to noise cancellation may also be adopted. One benefit of applying a multiple channel noise canceller is that the noise canceller can be used to cancel more than one noise signal or to enhance multiple signals. In milling machines, an applied cutter normally has numerous teeth. Examples are end mill, V cutter and an inserted tooth cutter. During machining, more than one tooth engages the workpiece at any one time. Hence the cutting signals are actually generated by numerous cutting teeth.

In this application a multivariable noise canceller can be used to detect cutting signals from multiple teeth and to cancel several noise signals. In practical machining, the milling machine may have only one noise source but the noise may have several vibration paths to the accelerometer. Thus the single source appears to be a multiple noise source due to these different paths. The multivariable noise canceller can also provide a significant noise reduction in this situation.

4. EXPERIMENTAL RESULTS

Figure 2a and 2b show the original measured cutting signal and the enhanced signal after processing using update equations (7) & (8). A series of ten different cutting experiments were carried out, five were carried out with a new tool and the remaining five used a blunt tool. The figure of the cutting signal indicates the periodic nature of the cutting signature. This has been described as the outcome of the mounted cutting tool being unbalanced. After a period of time of continual cutting this will cause uneven rates of wear on the cutter teeth.

The average signal to noise ratio (SNR) improvement in dB for the two-channel case considered in this experiment compared favorably with the ANC method outlined in [2]. The average SNR over both worn and new tool conditions was 4.5 dB.

5. CONCLUSIONS AND DISSCUSSION

It is interesting to note that [2] reports a significant increase in SNR improvement when four channels are employed in the noise cancellation system. This can be attributed to the additional information that is available to infer the transfer function parameters within the maximum likelihood estimation framework. The ANC method is based on a least mean squares (LMS) minimization technique that, of course, is the MLE under Gaussian assumptions. The approach proposed in this paper to signal enhancement overcomes the problem of requiring noise alone periods for filter adaptation. This blind approach utilizes some prior knowledge of the signal and noise statistics. The signal enhancement in terms of SNR is of comparable performance to the MVANC technique detailed in [2].

It is considered further work to investigate multi-variable extensions to the dual-channel model investigated here. The main motivation for the signal enhancement is a signal pre-processor for a dynamic classification machine for condition monitoring. It is proposed that the blind filtering model may be extended to accommodate adaptive density parameterization using for example mixture of Gaussian models. This may then provide a means of learning the class-conditional densities required for a statistical classifier.

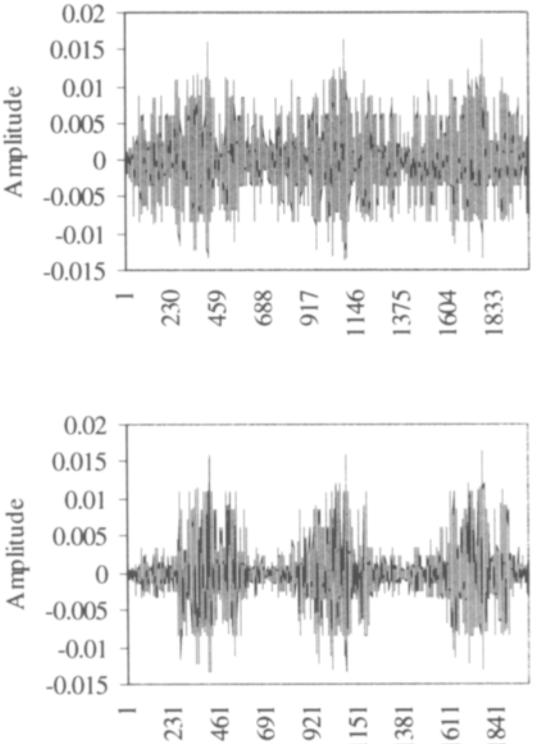


Figure 2a : Corrupted signal , **b** : Enhanced signal

REFERENCES

1. Widrow, B., Adaptive Noise Cancelling : Principles and Applications. *Proceedings of the I.E.E.E*, 63 (12), pp 1692 – 1716, 1975.

2. Hung, C, P., Moir, T, J., and Findlay, J., Multivariable Adaptive Noise Cancellation Applied to a Milling Tool Condition Monitoring System. *Applied Signal Processing*, 1:181 – 187, 1994.

3. Lee, T, W., Girolami, M., Bell, A, J. and Sejnowski, T. A Unifying Information Theoretic Framework for Independent Component Analysis. In Press *International Journal on Mathematical and Computer Modelling*, 1998.

4. Girolami, M., Symmetric Adaptive Maximum Likelihood Estimation for Noise Cancellation and Signal Separation. *Electronics Letters,* Vol, 33, No.17, pp 1437 – 1438, 1997.

5. Girolami, M. An Alternative Perspective on Adaptive Independent Component Analysis Algorithms. In Press *Neural Computation*, 1998.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Autonomous System For Oil Pipelines Inspection

Jun Okamoto Jr.[a], Julio C. Adamowski[a], Marcos S. G. Tsuzuki[a], Flávio Buiochi[a], Claudio S. Camerini[b]

[a] Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia Mecânica,
Av. Prof. Mello Moraes, 2231 CEP 05508-900 São Paulo, SP Brazil

[b] CENPES/PETROBRAS, Cidade Universitária, Quadra 7, CEP 21949-900 Rio de Janeiro, RJ Brazil

Maintenance of oil pipelines is an issue of great concern for oil companies. Any possibility of leakage must be detected before the leakage occurs and a preventive action should be taken in order to avoid losses of oil and ecological disasters. One of the main causes of oil pipelines leakage is the corrosion of the bottom part of the pipeline due to accumulation of water and other corrosive substances.

One of the methods used to check the conditions of the oil pipelines is the running of a data acquisition device through all the length of the pipeline (that can be kilometers) to gather information about the corrosion and its position inside the pipeline. This device is commonly referred to as *pig*.

The Brazilian oil company, PETROBRÁS, wanting to have the technology of such device to detect corrosion in their oil pipelines proposed to the Department of Mechanical Engineering of the Escola Politécnica da Universidade de São Paulo a joint project to the development of a *pig* that could run uninterruptedly in long oil pipelines gathering corrosion data for the preventive maintenance of their oil pipelines.

The result of the project was the development of an ultrasonic *pig* with 16 ultrasonic transducers with on-board energy system and acquisition and storage systems. Also, comprised in the presented solution was the development of a software to analyze the collected data and give the position of the corrosion spots along the pipeline.

This paper presents in detail the implementation and design issues related to the development of the ultrasonic *pig*. Also, some experimental data will be shown as confirmation of the effectiveness of the developed system.

1. INTRODUCTION

Oil pipelines are built with carbon steel pipes, with diameter ranging from 200 mm to 500 mm, that can be extended for hundreds of kilometers uninterruptedly, for carrying petroleum and its derivatives [4]. Those pipes are prone to internal corrosion usually caused by presence of water (salty or not), that causes material loses of the pipe wall in the bottom part of the pipe.

In the case of oil pipelines inspection, that usually have great extension and restricted access, there is a need for an autonomous inspection equipment that can run through inside of the pipeline measuring characteristics of the pipeline wall.

For each type of corrosion there is an adequate inspection technique, that can vary from visual inspection, X-ray, Eddy currents, acoustics and ultrasonics [1]. Magnetic and ultrasonic [3] methods are the most commonly used to detect this loss of material. The magnetic method consists in the application of a magnetic field in the internal wall of the pipe using permanent magnets. The distortion of the applied magnetic field due to the change in the pipe wall thickness is detected by magnetic sensors. This method has been applied to detect defects in the pipe inner wall as well as in the outer wall. The measurements obtained with this method are only of qualitative nature. By the other hand, the ultrasonic method allows the obtaining of quantitative results such as the extension and depth of the defect with the precision of millimeters [5]. This method is based on the measurement of the propagation delay time of an acoustic wave, emitted by an ultrasonic transducer, that travels in the liquid medium and that is reflected when the wave reaches the inner and outer wall of the pipeline. The distance between the transducer and the pipe wall can be obtained from the signal reflected by the pipe inner wall, and is used to evaluate the internal corrosion. The time of flight between the echoes from the inner and outer wall is related to the pipe thickness and allows the evaluation of the external corrosion. The dimension of the defects to be detected depends on the characteristics of the ultrasonic transducer, the precision of the time measurement of the propagation of the acoustic wave and the synchronization of the transducer's trigger with the measurement of the *pig*'s movement.

2. GENERAL DESCRIPTION OF THE *PIG*

The mechanical part is composed by a cylindrical capsule supported by two rubber discs. The capsule is allowed to rotate freely around its longitudinal axis and has the center of gravity under this axis to keep the ultrasonic transducers always directed to the pipeline's lower half. The rubber discs also keep the capsule centered in the pipeline and serve the purpose of blocking the fluid to propel the *pig*. Held outside the capsule is an odometer wheel to measure the traveled distance of the *pig* inside the pipeline. The odometer wheel keeps permanent contact to the pipeline's inner wall and generates electrical pulses for each 100 mm of traveled distance.

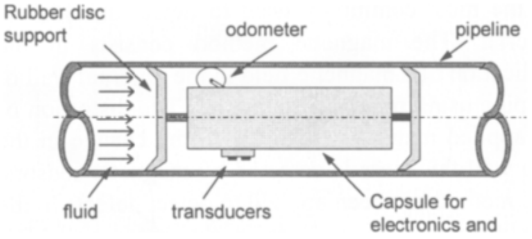


Figure 1. Schematics drawing of the *pig* inside the pipeline

Inside the capsule are the electronics and batteries of the *pig*. The electronics assembly is composed of 5 printed circuit boards, with 80 x 100 mm, in a piggyback mount, with 61 mm of total height. The 5 printed circuit boards are: hard disk, microcomputer, power supplies and echo time of flight detectors and 2 board for transducers excitation and reception of echoes with 8 channels in each board. This set of boards is fixed to the capsule's plug by rubber suspensions to avoid excessive vibration on the boards. Also, in this plug are fixed the connectors for external connections (communications and odometer) and transducers.

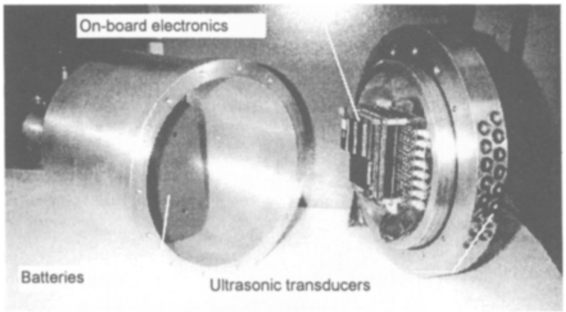


Figure 2. *Pig* prototype with internal elements

The area to be inspected in the lower part of the pipeline corresponds to a stripe of 150 mm in length. The *pig* uses 16 ultrasonic transducers mounted in a mechanical support in the capsule's plug. The mechanical support and the rubber discs keep the transducers 30 mm away from the pipeline inner wall. The transducers are used in pulse-echo mode to measure the echo time propagation. The time measured is converted to distance considering that the speed of the sound propagation in the fluid is known (petroleum and its derivatives).

In the beginning and in the end of an inspection a microcomputer, IBM PC type, is connected, via serial link, to the *pig*, for the programming of the inspection and data retrieval. Before the inspection the microcomputer connected to the *pig* is used for calibration procedures.

The fluid contained in the pipeline serve as coupling between the transducers and the pipeline's inner wall. The speed of propagation of an acoustic wave varies with the fluid and also with the temperature, so the fluid in the pipeline during the calibration procedure is the same that will be used during the inspection.

After the calibration process, the *pig* is kept in a sleeping state, waiting for the inspection to begin.

The beginning of the inspection is detected by the generation of pulses by the odometer wheel. The odometer has a resolution of 100 mm and its circuitry is capable of generating 5 to 10 pulses per second depending on the speed of the *pig*. (it is designed speeds from 0.5 m/s to 1 m/s). If the *pig* does not detect pulses from the odometer wheel for more than 1.0 second, the data acquisition is interrupted assuming that the *pig* has stopped, and the *pig* enter a sleeping state, resuming the inspection after pulses are detected again. This procedure is necessary for optimizing the energy consumption.

A typical measurement rate for the *pig* is around 250 measurements per second per transducer, although all the circuitry has been developed for coping with a rate of 500 measurement per second per transducer. Considering the rate of 250 measurements per second and a traveling speed of 0.5 m/s, the *pig* executes one measurement every 2 mm of distance change. The defects (corrosion spots) usually extend for more that 10 mm. Considering the 16 transducers of the *pig*, the total rate of data acquisition is around 4,000 measurements per second.

The software in the *pig* classifies the received echoes in intervals related to the defect depth. This classification is based on the percentage of the wall corrosion, in discrete intervals. In this version, were used 8 levels of classification according to the corrosion depth. The 8 level was chosen to save storage space, so that each transducer channel reading could be stored in a 3-bit variable.

During data acquisition, the software uses data compression mechanism in order to save storage space. The compression mechanism is based on the fact that is not necessary to store information when there is no defect in the pipeline. When a defect is found in during the *pig*'s movement, the level of corrosion and its position are stored. A pipeline usually has 5 to 10% of defects in its whole extension.

The transducers in the *pig* have 5 MHz of operating frequency and are focused by acoustic lenses of polymetil-metacrilate (Lucite). Also, these lenses serve the purpose of protecting the transducers from the high pressure. The *pig* is designed to stand up to 100 ATM and operation temperature from 5 to 45 °C.

3. ON-BOARD ELECTRONICS

The choice of the control system for the sensors and data storage system was done considering the operation with batteries for long periods of time. Therefore, the on-board electronics have characteristics of low power consumption and high capacity of data storage. Additionally, due to the variation of the battery's tension along the time, the specified system had to operate in a wide range of power supply voltage. The Figure 3 shows a block diagram that represents all the on-board electronics.

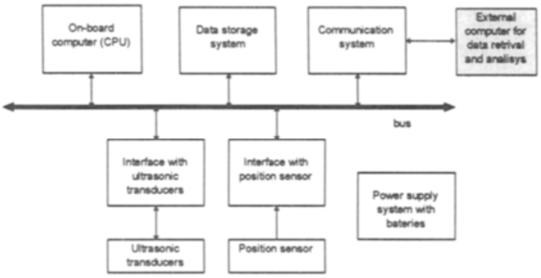

On-board computer (CPU)
Data storage system
Communication system
External computer for data retrival and analisys
bus
Interface with ultrasonic transducers
Interface with position sensor
Power supply system with bateries
Ultrasonic transducers
Position sensor

Figure 3. Block Diagram of the on-board computer

The CPU board is based on the Hitachi H8/532, a 16-bit microcontroller, operates at the clock frequency of approximately 20 MHz, 45 kbytes of program memory, 512 kbytes of data memory, and has specification of low power consumption. This board has an internal multitask system for developing programs in Forth and assembler.

The 512 kbytes of memory is used as temporary data buffer for the collected data to be sent to the hard disk unit, that can be completely filled with corrosion data.

Additionally, are available 26 to 41 bits of I/O ports, depending of the configuration programmed in the CPU, 8 multiplexed channels of an 10-bit A/D, 3 channels of 8-bits D/A and 2 serial communication channels used for communication with an external computer and for extraction of the data from the storage system.

The supply voltage supported by the CPU board ranges from +6V to +16V, with a typical current consumption of 31 mA and a 300 μA of stand by current consumption.

The secondary storage system is composed of a micro disk with 40Mbytes of data storage capacity, mounted in piggyback with the CPU board. The micro disk is manufactured by Citizen, in glass, with 1.3 inches of diameter.

The interface with the transducers is done by two transmitter boards and one receiver board. Each transmitter board is responsible for exciting 8 transducers in multiplexed mode. The excitation of each transducer is done by a narrow pulse, with 50 ns of duration, and 200V, through a MOSFET driver. The reception is done by pre-amplifying the echo with a protection circuitry for the input. The pre-amplified echo passes through another amplifier circuit to an echo detection circuit, based on a fast comparator with response time less than 50 ns. The echo time of flight is measured by an 11-bit counter with 20 MHz clock. The time measurement allows a resolution of approximately 0.04 mm in the calculus of the defect's depth. This board also includes the circuitry for the odometer, power supply connection and generates interruption for the CPU board.

The *pig*'s power supply is composed by a set of Nickel-Cadmium batteries with nominal tension of 12V and a capacity of 7 Ah. The total consumption of the whole system was estimated in 50.4 Wh.

Additionally, is necessary an external system composed of a microcomputer, with data manipulation software and a battery charger unit.

4. ULTRASONIC TRANSDUCERS

The ultrasonic transducer was chosen based on the axial resolution and the acoustic attenuation allowed for the application. The greater the transducer frequency the wave length is shorter and therefore the better is the axial resolution. However, the acoustic attenuation in the fluid increases exponentially with the frequency increase. So, as a compromise, was selected a transducer that operates in 5 MHz as a solution for this application.

The ultrasonic transducer used in the *pig* is a wide band type with central frequency of 5 MHz and 6 mm of diameter. Without a focusing lens this transducer has a beam width of 3.5 mm at a distance of 50 mm. In order to improve the axial resolution, that is related to the beam width, an acoustic lens with focal distance of 35 mm and focus width of 2.0 mm was used. The Figure shows the mapping of the acoustic field produced by the transducer at the frequency of 5 MHz. In this figure the transducer and lens is positioned to the left.

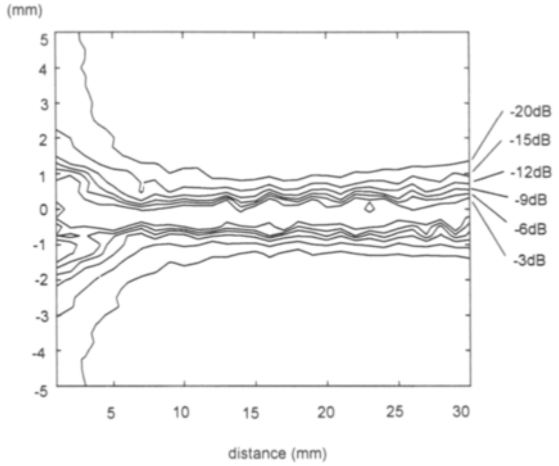


Figure 4. Acoustic field map for the ultrasonic transducer at a frequency of 5 MHz

5. CONTROL AND DATA ACQUISITION

5.1 Measurement Control

The *pig*'s on-board computer controls 16 transducers, activating one at a time and executing the measurement of the pulse time of flight in order to determine the occurrence of corrosion or not in the pipeline's inner wall.

Two internal counters of the CPU board are used to generate two types of interrupts: (1) real time interrupt with 10 ms interval; (2) transducer's trigger interrupt, variable interval depending on the *pig*'s speed. The second interrupt can have intervals of 100 μs to 1.6 ms depending on the *pig*'s speed to maintain the data acquisition constant in relation to the *pig*'s movement. That is, the transducers are triggered for a new measurement cycle every 2.0 mm.

The odometer wheel generates one pulse each 100 mm through a set of permanent magnets and reed switches. The reading is done by one bit of the parallel ports at each real time interrupt (every 10 ms). Considering the odometer's resolution, at the *pig*'s maximum speed of 1 m/s, a pulse is detected each 100 ms. A new speed value is calculated every second, if the speed is above 0.1 m/s and every 5 seconds if the speed is below 0.1 m/s. The *pig* is considered to be stopped if the calculated speed is below 0.02 m/s.

5.2 Defects Classification

The detected defects, based on the time of flight of the ultrasonic pulse, are classified in 6 levels of defect and 2 additional levels that represent "no defect" and "echo lost". Therefore, to classify one defect is necessary a 3-bit word. The level 0 means that the pipeline is in good condition, this level is assigned to a transducer when the time of flight is found between a reference value and the reference value plus a programmable offset between 1 and 6 mm. The levels of defect from 1 to 6 are assigned to a transducer when the calculated depth is above the offset value with programmed increments of 1 to 3 mm between each level. The level 7 is assigned to a transducer when the time of flight is above a limit value, in this case the echo is considered to be lost.

5.3 Data Storage

The collected data cannot be integrally stored in the hard disk due to the disk's storage space limitation. Consequently, only the information about the defects and their location is stored in the disk. The data storage is controlled by a program that allows the recording of sequential data as the disk were a magnetic tape, which is convenient because of the measurement characteristics of the *pig*.

In the case of a defect detection, the odometer reading and the defect level are stored in the disk. Values corresponding to the level 0 are not stored.

When the *pig* is considered to be stopped by the odometer reading the data are no longer stored, even if a defect is being detected.

5.4 Operating Procedures

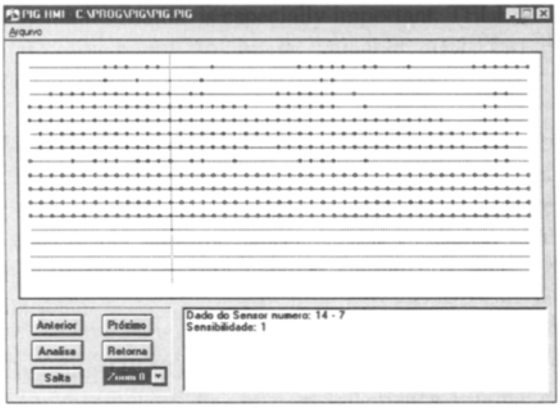
The connection of the *pig*'s capsule with the external world is done by two high pressure resistant connectors. One of the connectors is for the power control, including battery charge, battery load measurement and shut off the power supply for the *pig* when not in operation. The other connector is for communication purposes. The serial channel is used to test the measurement functions, program the parameter for an inspection and retrieving the data after the inspection is completed.

When the power is first applied to the *pig* the on-board computer enters a state for waiting for a command to come through the serial channel. Basically, 4 commands can are recognized: measurement functions test, programming of defect's levels, beginning and ending of inspection. After the begin inspection command is issued, the *pig* waits for the movement start looking for odometer pulses.

Once the inspection is started, the *pig* will not stop collecting data until an end of inspection command is issued. After this command the data stored in the internal hard disk can be transferred to an external computer for analysis.

6. DATA VISUALIZATION

To visualize the data collected by the *pig* a software written in Borland C++ for Windows was developed. The software interprets the coded data retrieved from the *pig* and displays it in graphical form where the whole extension of the pipeline is showed and all the detected defects are represented for each transducer. The software also allows zoom in the data to visualize details of the defect groups.



Dado do Sensor numero: 14 - 7
Sensibilidade: 1

Figure 5. Screen shot of the data analysis program

7. EXPERIMENTAL RESULTS

The *pig* was tested both in laboratory and in the field. In the laboratory tests corrosion spots were artificially made in a pipeline segment and the data was exhaustively stored to determine the effectiveness of the system. Figure 6 shows the results for the collected data where the depths are in millimeters demonstrating that the readings of the sensors give a very precise depth estimation.

The axial resolution of the ultrasonic system is 0.04 mm for water and is dependent on the sound propagation speed in the liquid. Petroleum has a coefficient of acoustical attenuation much higher than water which required a high gain amplification of the input signal, around 60 dB. When operating the *pig* in water the axial resolution is reduced due to the excessive gain of the amplifier.

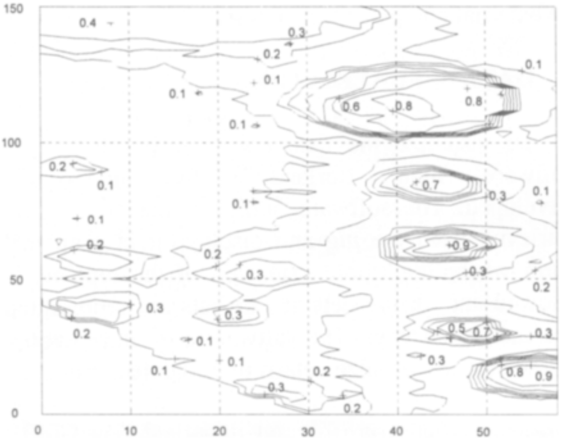


Figure 6. Round portions are corrosion spots detected during an inspection. All numbers are in millimeters.

Field tests showed that the procedures for calibrating the measurements, starting and stopping the inspection, data retrieval and analysis and battery energy utilization were adequate for the developed equipment.

8. CONCLUSION

The development of an autonomous system for oil pipelines inspection was described in this paper. The results presented showed the practical application of the developed equipment.

The on-board electronics resulted very compact and with low power consumption. Also the data storage system, based on a micro hard disk, resulted in a very robust solution even during field tests. The transducers used in the *pig* allowed the detection of corrosion spots with 10 mm in diameter and 1 mm in depth with very good repetitibility in only one bit, representing 0.04 mm in depth.

More field tests will be conducted in the future and more processing power and data storage capacity will be added in the on-board electronics. This will allow more exhaustive data collection and complex data analysis.

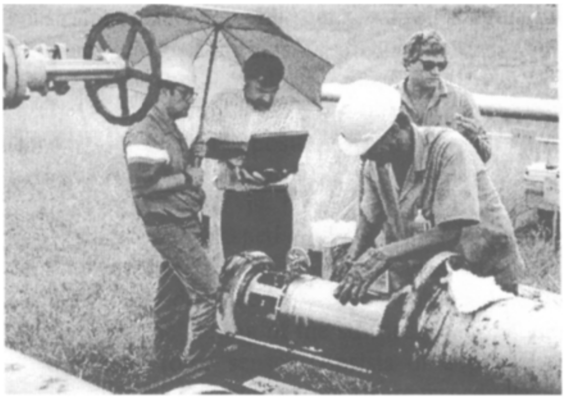


Figure 7. *Pig* being introduced in the pipeline for field testing

ACKNOWLEDGEMENTS

This project was a supported by Petrobás (Brazilian Oil Company) and developed at the Department of Mechanical Engineering of the Escola Politécnica of the University of São Paulo. The presentation of this paper was supported by CAPES.

REFERENCES

1. Hagemaier, D. J.; Wandelbo, A. H.; Bar-Cohen, Y., "Aircraft Corrosion and Detection Methods", *Materials Evaluation,* 43(4), p. 426-437, October 1984.
2. "Pipeline Testing with Pigs", Das Echo (The Echo) - Information on Testing Materials with Ultrasonics, n. 34, p. 12-13, 1989.
3. Goedecke, H. "Corrosion Surveys with the UltraScan *Pig*", Corrosion Prevention and Control, p. 33-36, April 1990.
4. Weisweiler, F. J.; Sergeev, G. N., Non-Destructive Testing of Large-Diameter Pipe for Oil and Gas Transmission Lines, Weinheim, VCH, 1987.
5. Edelmann, X.; Gribi, M., "Mechanized Piping System Testing by Ultrasonics and its Contributions to Safety and Economic Performance", Sulzer Technical Review, p. 48-50, 1990.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Novel 3D Input Device, with 6 DoF, Prototype Based Mechanically on a Gyro and Electronically via an FPGA

Peter E Jones

Electrical and Electronic Engineering, The University of Western Australia,
NEDLANDS WA 6907, Western Australia

A novel input device for use with desktop computers that has six degrees of freedom (6 DoF) is prototyped. It is a three-dimensional (3D) input device. Mechanically, the inspiration was from the gimbal mount of a gyroscope. A Field Programmable Gate Array (FPGA) was used together with a suitable programming environment to rapidly develop and alter the electronics. Interactive systems that use a 3D world should ideally be complimented with a 3D input device. Such devices have previously been built mainly for immersive virtual reality (VR) systems and are not well suited for desktop systems. This device is intended for use with the non-dominant hand, freeing the other hand for finer-motor skills to use a conventional mouse or the keyboard. The aims of being ergonomic and providing a good cognitive match for navigating in 3D worlds have been met.

1. INTRODUCTION

Virtual Reality's software and technology is now recognised for its value in emerging applications that use three dimensions (3D). This novel 3D input device (Jones & Garcia-Webb 1997) with 6 degrees of freedom (DoF) is based mechanically on a gyroscope's gimbal mounting and uses a Field Programmable Gate Array (FPGA) to prototype its electronics. We have named it 3DGyro to acknowledge this. Effortless hand-eye co-ordination is best provided by a device that takes advantage of the human visual and proprioceptive abilities. The lengthy and dominant reign of the 2D mouse is at an end when it comes to 3D systems.

The applications for a 6 DoF input device range over new user interfaces, web browsers, computer-aided design & manufacture, computer simulation, virtual reality, telerobotics, scientific data visualisation, molecular modelling and other interactive computer graphics.

2. AIMS FOR 3DGYRO

We wanted a device that would allow a user, in a non-immersive 3D world, to fluently move through 3D; pick up objects and move with them; or not move and just manipulate the object. At the same time as performing these tasks with the non-dominant hand, we wanted the dominant hand to carry out finer-motor controlled manipulations with a standard mouse. Work elsewhere has shown the advantages of two-handed input (Cutler et al, 1997).

For the device itself to be practical it needs to be low cost, after all the standard mouse has tumbled to a price that is so low it is pointless producing a device that would be orders of magnitude more expensive. This can be achieved by using off-the-shelf components as far as practical. Most users place a premium on their desk space so it needs to be compact. As it would be in use through much of the working day it should not be tiring to use and it must be intuitive to use. We also did not want to tie the user to the device with either cords or any attachments. The electronics needed to be easily modifiable as changes or constraints appeared, and for this reason we went for a gate-array prototyping system. This was expensive and would not be needed for any later versions. From a software point of view it needs to be useable with most 3D applications. Then it needs to interface to most computers over a serial interface. Any new device with additional functionality can be difficult to retrofit to existing applications unless they are written with adaptability in mind, such as the kernel in Maverick (Hubbold et al, 1996). Finally, the device needs to be robust, reliable, and easy to maintain.

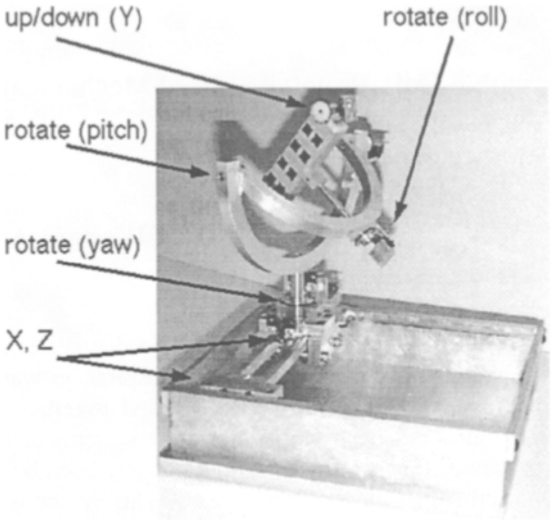


Figure 1 — 3DGyro Prototype

3. INPUT DEVICES

It is possible to categorise (Jacob, 1996) input devices into type of motion (linear vs. rotary); absolute or relative measurement; physical property sensed; number of dimensions; direct vs. indirect control; position vs. rate control; integral vs. separable dimensions (e.g. mouse and Etch-a-Sketch). Jacob emphasised the need to allow users natural and expressive operations. Transducers that are used either measure displacement or strain. Displacement transducers can be based on electrical or magnetic properties (resistance, capacitance or inductance), or one of several optical techniques including the use of TV cameras. Those using strain gauges can be moveable or fixed. Some devices give feedback to allow users to sense objects. A number of useful references to input devices for 3D are in Hand's publications (1994 & 1997). See the URLs for additional input devices such as mice, balls, gloves and haptics.

Mice operate in relative mode (positional change relative to current position and orientation). For 3DGyro we decided to also have parametric (or rate), where the speed of movement in the display depends on device positioning.

There are several variations of the mouse that allow the user to input and manipulate data in three dimensions. Some use co-operative action of the two hands and the proprioceptive information. The Logitech Magellan (see URLs) mouse provides 6 DoF. Finger pressure is used for small deflections in x, y and z, twists cause rotation about an axis.

Polhemus (see URLs) are a manufacturer of electromagnetic 3D measurement systems. SpaceOrb™ 360 (see URL for Spactec) is aimed at 3D action games and uses a six-axis ball providing 10-bit resolution with several buttons. It is meant to be used with two hands.

Human movement tracking can use sensors and sources on the human or have just sensors on the body detecting outside sources, or an external sensor that electro-optically tracks the human. One tracking system is based on the user wearing an instrumented glove. The 3D tracking devices provide gloves with absolute x, y, z, yaw, pitch, roll position and orientation information.

Haptics Community Web page (see URLs) has a set of pointers to work involving physical feedback. At MIT there is also research into haptics (see URLs). They are investigating how controlled alterations in visual, auditory, and haptic displays affect human perception. Their results are being used to overcome some of the technological limitations and are applicable to the design of optimal human-machine interaction paradigms.

The PHANToM (see URLs) allows the user to interact with the computer by inserting his or her finger into a thimble. With it, users are able to touch and manipulate virtual objects with haptic feedback. It provides 3 DoF for force feedback, and optionally, 3 additional DoF for measurement.

Commercial haptic systems (see URLs) also have been made available mainly for medical applications, although lower priced joystick versions exist for the games market. They also produce 3D digitisers, which could allow 3D navigation too.

3.1. Comparisons

A series of five experiments at the University of Toronto (Zhai, 1995) have been carried out which confirm the general direction we have taken. The devices compared were: a Spaceball, EGG (Elastic General purpose Grip — a tethered rod), Fball (rather like the Polhemus 3Ball) a ball with Ascension Bird in it, and a locally developed MITS (Manipulation In Three Space) glove.

Except for the games-based devices, all these devices are expensive; use too much space; are uncomfortable, tiring, and difficult to use; poorly matched to the tasks in non-immersive VR and often need attachments that entrap the user.

4. DEVELOPMENT OF 3DGYRO

We decided to make 3DGyro a desktop device to allow support of the user's arm, for example by resting the elbow on the desk. A desktop device had the advantage of maintaining deskspace, avoiding the risk of it falling off the desk and allowed it to be used in either of absolute, relative or parametric modes.

Mechanically, 3DGyro is based on something that looks like a gyroscope, as shown in Figure 1. That is, it is like an open globe that you place your hand in and grip a handle. This handgrip is gimbal mounted and you can tilt it (pitch), twist it (yaw) and rotate (roll) as much as the movement in your wrist allows. We considered measuring the downward pressure for the up/down movement (Y), but for now, have settled on a thumb-wheel design placed at the top of the handgrip. The whole assembly can be moved as if it were a 2D mouse but is on a constrained platform base (not unlike a flatbed pen-plotter). This provides the X and Z movement. Underneath the finger positions on the handgrip are eight buttons. The sensors for each of the three shafts on the gimbal mount, the Y thumb-wheel, and the X, Z optically measure the rotations/translations using low cost shaft encoders (as done in many mice).

To allow for both rotational and translation the gyro-like device is mounted above a 2D translation mechanism. The user can then provide both translations and rotations.

Electronically, the FPGA is a Xilinx FPGA 401E with 576 Configurable Logic Blocks mounted on a prototyping board marketed by Virtual Computer Corp. This is an ISA PC bus card. From 3DGyro are the inputs from the eight buttons, five quadrature waveform pairs from the shaft optical encoders and one further pair from the thumb-wheel. The only output to 3DGyro is power for the optoelectronic sensors and for the switches. This use of a re-programmable FPGA gave us flexibility at the hardware level to accommodate changes in the design. After completing the design it was possible to simulate the operation of the system using a software test bench.

We have begun with software written by David Stampe when he was at the University of Toronto, Canada — VR386. Other researchers have offered us access to their systems including the source so that we can change the way that input events are dealt with. However, for the initial evaluation we found that VR386 is sufficient. Surprisingly, it needed few changes to accommodate a device that was not available when the software was designed. VR386 provides a 3D world, populated with objects. It allowed us to evaluate 3DGyro in terms of ease of navigation, picking up and manipulating objects. At the same time VR386 allowed the use of a conventional mouse. The prototype was made out of aluminium, which is strong, light and easy to work. It is likely that plastic could be used in later versions.

The sensitivity-up/down control increments or decrements a two-bit counter in the hardware and this can be read by the software to adjust the sensitivity. For coarse navigation a small displacement would produce a large movement if the sensitivity was high and vice versa for when finer movements are required.

The most convenient locations for the opto-sensors would be at the pivot points of the shafts. This direct mounting would have limited the resolution to the number of signal changes possible from the spokes in the light beam interrupter. However, with use of gearing the beam interrupter can then provide several times this resolution. The wrist limits the range to a realistic movement of around ±60°. Without gearing this would mean a resolution of 50 clicks or about 1°, however, with a gearing of 5:1 this can be increased to 250 clicks, which is acceptable.

4.1. Translational

The horizontal translations correspond to the X and Z movement and the vertical is the Y. That is a left-right movement is ±X and forward-back movement is ±Z. Our early thoughts were that the whole device would be like some giant mouse and move around the desktop. The user's reach and the free space on the desktop would then control the range.

The idea was rejected on several grounds. There would be no intuitive zero position for the parametric mode. Few users maintain sufficient clear desk space. It would be too easy to drop the device off the edge of the desk or collide with another object on the desk. Either would result in possible damage to the device. Ordinary mice are smaller, lighter and usually are protected from collisions by the user's hand. If a mouse falls off the desk it is usually not a problem — they are light enough to dangle on their lead. They are also cheap enough to completely replace them even if they are damaged. The 3DGyro would also have the problem that rotations might move the whole device, giving rise to unwanted translations.

We next considered a desktop box of similar size to a mouse pad. It would have two perpendicular sliders, one for X and one for Z. The sliders would be strong enough to support the weight and forces of the gimbal mount. The translations are converted to rotations using the same opto-sensors as used for the rotations. We decided on a higher resolution for the translation of 500 clicks. In a slider of length 15cm and a 40-slot disk, with a design that obtains the maximum resolution for each transition, meant that it required a gearing such that it rotated once per 5cm. That is 15cm allows 3 revolutions × 40 slots × 4 transitions giving 480 clicks.

The final version used a full-length slider for the X-axis and the Z-axis slider was shortened so that it slid along the full-length of the X-axis slider rather than being right across the base support. This not only reduced the mass but also improved the movement of the slider itself.

4.2. Electrical Interface

The Xilinx FPGA came mounted on a prototyping board marketed by Virtual Computer Corp. This used the Industry Standard Architecture (ISA) bus for IBM compatible PCs (see Figure 2). Although expensive, at about AUD$1 200, it allowed for rapid changes to the design of the electronic interface.

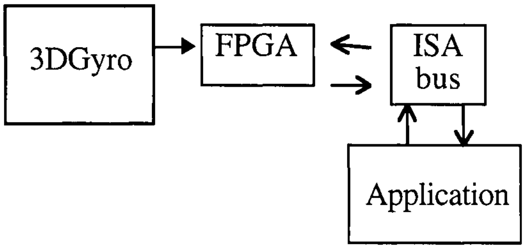


Figure 2 — Hardware Architecture

FPGAs are programmable logic chips (Gopakumaran, 1997). They allow rapid prototyping of digital logic and can be re-programmed as many times as you need even after installation. FPGAs are a mixture of logic cells (Configurable Logic Blocks CLBs) plus cells for the input and output of data (Input/Output Blocks IOBs). The board we used has a Xilinx 401E with 576 CLBs. Either schematic entry or a hardware description language (HDL) can specify the logic design. We used the Very High Speed Integrated Circuit (VHSIC) HDL or VHDL. It is an IEEE standard (1076-1993). After completing the design it is possible to simulate the result using a software test bench. We made some simplifications in our test cases as regards to some of the timings. The next step is one of synthesis, which transforms the design to a bit stream for the FPGA. This is provided as a header file that can be incorporated in a C program. The bit stream is then downloaded over the ISA bus during the initialisation phase of the application as it starts up.

The system is not reset on power up as it waits until the application is run which then downloads the bit-stream to the FPGA to configure it. To avoid the need to provide for a software initiated reset of the internals of the FPGA it was decided to provide an external reset by simultaneously pressing the clutch and third user button (both in the second column under the index and middle finger). To have had a software reset would have meant writing to the FPGA whereas the current design provides only for the software reading from it. It would have added complexity to the interface for very little gain.

The opto-sensors are in pairs and provide a rectangular digital waveform that is in quadrature phase. The interface needed to determine that a change in one of the input signals had occurred and from that phase information what the direction was. Having obtained the fact that there is a change and its direction we decided to use a 16-bit up/down counter.

The clutch button had choices for its design. For incremental mode its use was to allow the user to reposition the handgrip without altering the display position. It could be either a full hardware implementation or a simple signal to the application. As we were using 16-bit counters the use of a clutch could lead to the counters wrapping round, i.e. either overflowing or underflowing. For parametric mode the clutch button reset the counters to zero, in a similar way to them being cleared when switching to this mode.

The user buttons (U1, U2 & U3) are simple on/off signals readable by the application. For the single degree of freedom button (1 DoF) there was not enough logic left in the FPGA to implement it in hardware.

The FPGA design used 534 CLBs (92%). Actual resolution achieved was 500 ticks for X & Z, 300 for rotations.

The application used for initial testing was VR-386 from David Stampe developed at the University of Toronto following on from REND386. We made changes to its initialisation, expanded its data structures to cope with the additional features of 3DGyro and scaling for 6 DoF. The user buttons were allocated as follows. U1 to pick up an object, U2 to manipulate an object without moving the virtual user, U3 separated rotation and translation (either all translations relative to rotational orientation, or relative to virtual world axes). At the same time the standard mouse was used to select objects whilst navigating using 3DGyro.

5. EVALUATION

The prototype used low cost parts amounting to about AUD$200 but there was much uncosted work in the workshops. It used off-the-shelf components, it was compact, reasonably robust, and using the FPGA allowed ease of modification. Of course, the expensive FPGA board would not be required for any later versions it would use low cost ICs.

Initial experiments with 3DGyro has shown it to be a better cognitive match than using a 2D mouse with meta keys (Jacob & Oliver 1995) or in using other techniques such as the virtual sphere. We have found it easy to use with the non-dominant hand. The dominant hand, meanwhile, is left free to control an ordinary 2D mouse to carry out simple selections and finer control via means of control handles on selected objects. It requires mechanical counter-balancing and improved shaft encoders.

Trials with several users operating it with VR386 worlds showed that 3DGyro was not tiring to use, and all users claimed it was intuitive in use. We were pleased to see how easy it proved to use with the non-dominant hand and that it was easy to use the mouse in the dominant hand at the same time. With its prototype limitations it was only possible to use it with right-handed users' non-dominant hand. Other research has shown that for various reasons left-handed users would not find it a problem to use (as they reluctantly have needed to become skilled at using either hand), but we would prefer to cater for both types of users.

There are problems beyond those that were deliberately not addressed in the prototype. There is some slippage between the axes and the driving of the slotted disks. The mechanism needs to be spring loaded to maintain reasonable pressure due to tolerances and to allow for wear or accumulated dust. It assumes that the non-dominant hand is the left one, the handgrip needs a thumbwheel at the top and bottom in order to allow for it to be fully rotated and so be useable by either hand. The buttons need to be easier to press and to give improved sensory feedback when pressed. The gimbal mount needs some counterbalancing to allow the user to take their hand away and have it remain stationary. We still don't see that springs are the answer to this and are looking at other means such as simple counterweights to balance the mechanism.

It is a pleasure to acknowledge the efforts of David Garcia-Webb, the staff in the general workshop and the electronic workshop in not only creating the working prototype but helping with its design. Thanks also to Gary Bundell and John Morris for their assistance with the FPGA software used in the implementation. Its development was internally funded

REFERENCES

Cutler, L.D., Frölich, B. and Hanrahan, P. (1997), "Two-handed Direct Manipulation on the Responsive Workbench" ACM Proceedings of the 1997 symposium on Interactive 3D graphics, pp. 107-114.

Gopakumaran, B. (1997), "Gate Array IC Design Tools". IEEE Computer Vol. 30, No. 1, pp. 117-118, 121-122, January 1997.

Hand, C. (1994), "A Survey of 3-D Input Devices", Technical Report CS TR94/2, Department of Computer Science, De Montfort University, Leicester, UK. [updated as TR96/7] Also at http://www.cms.dmu.ac.uk/People/cph/VRbib/Tracking/trackp.html

Hand, C. (1997), "A Survey of 3D Interaction Techniques", Computer Graphics Forum, 16(5): 269-281 (Dec 1997).

Hubbold, R.J., Dongbo, X., Gibson, S. (1996), "MAVERIK - the Manchester Virtual Environment Interface Kernel", Proceedings of 3rd Eurographics Workshop on Virtual Environments, Monte-Carlo, February, 1996, European Association for Computer Graphics, P.O. Box 16, 1288, Switzerland.

Jacob, R.J.K. (1996), "Human-Computer Interaction: Input Devices", ACM Computing Surveys, Vol. 28, No. 1 (March), pp. 177-179.

Jacob, I. and Oliver, J. (1995), "Evaluation of Techniques for Specifying 3D Rotations with a 2D Input Device", Proceedings of HCI '95 Conference, People and Computers X, ed. Kirby, Dix and Finlay, Cambridge University Press, pp. 63-76.

Jones, P.E. and Garcia-Webb, D. (1997), "Low-cost 3D Input Device for Next-Generation User Interface", Proceedings of First International Conference on Information, Communications & Signal Processing, Volume 3, IEEE Publications, ISBN 0-7803-3676-3, Singapore, 9-12th September, 1997, pp. 1672-1676.

Zhai, S. (1995), "Human Performance in Six Degree of Freedom Input Control", Ph.D. Thesis, Ergonomics in Teleoperation and Control Laboratory (ETC), University of Toronto. Also at http://vered.rose.utoronto.ca/people/shumin_dir/papers/PhD_Thesis/top_page.html

URLs (correct in May 1998)

Abrams/Gentile Entertainment, http://www.ageinc.com/age/

Ascension Technology Corporation, http://www.ascension-tech.com

Cosmo Player from SGI, http://cosmosoftware.com/

Creative Labs, http://www.creaf.com

General Reality Company, http://www.genreality.com

Haptics Community Web Page, http://haptic.mech.nwu.edu

Immersion Corp, http://www.immerse.com/WWWpages/impulse_engine.html

Lanier, J. (VR pioneer), http://www.well.com/user/jaron/index.html

Logitech, http://www.logitech.com

Microsoft, http://www.microsoft.com/VRML

MIT Touch Lab http://touchlab.mit.edu

Polhemus Inc., http://www.polhemus.com

SensAble Technologies, http://www.sensable.com

Silicon Graphics, http://cosmosoftware.com/galleries/

Spacetec, http://www.spacetec.com

Virtual Technologies, http://www.virtex.com

VR386 — David Stampe ftp://ftp.psych.utoronto.ca/pub/vr-386/

VRML players Repository at SDSC http://www.sdsc.edu/vrml/

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

The Application of a Systems Methodology to the Design and Specification of an Intelligent Telecare System

G Williams[a], D A Bradley[b] & K Doughty[a]

[a]SEECS, University of Wales,
Dean Street, Bangor, Gwynedd LL57 1UT, UK

[b]School of Engineering , University of Abertay; Dundee,
Bell Street, Dundee DD1 1HG, UK

Faced with increasing demands for resources for health and social services the introduction of advanced telecare systems is increasingly seen as a means by which technology could be used to support older members of society and the physically disadvantaged in living in their own home environment. In order to understand the needs of such a system and to identify the levels of integration and interaction involved a formal systems methodology has been used to provide a framework within which the individual components, many of which are inherently mechatronic in nature, can be encompassed.

1. INTRODUCTION

In order to provide high quality healthcare and to improve the quality of life of the elderly and the physically disadvantaged there is a need to provide a range of technologies and systems which will provide support for and access to facilities both in their home environment and elsewhere. This in turn requires an understanding of the role of individual technologies as well as the ability to integrate a range of advanced and mechatronic systems such as smart or intelligent wheelchairs for mobility, patient care and support systems and advanced dwelling designs leading to barrier free housing incorporating 'smart' or 'friendly' home technologies [1].

In particular, developments in information technology together with an increased understanding of the design requirements of complex mechatronic systems and of the associated safety implications of such systems now affords the opportunity to develop a range of advanced, intelligent and above all user centred systems which will provide enhanced support for independent living.

Indeed, telecare and 'hospital-at-home' services such as home emergency alert, community health information networks and ambulatory monitoring are becoming increasingly common [2-4]. However, many such systems, as well as others such as mobility aids, have tended to be developed in relative isolation, without due regard for the need for the provision of the open and integrated systems approach which is ultimately required to ensure that the necessary and appropriate levels of capability and capacity are in place.

All of the foregoing implies the design and provision of an intelligent system which is capable of receiving information from a number of dispersed sources, of analysing this information and of communicating both internally and externally. In order to achieve the requisite level of performance there is an obvious need to fully understand the role of each of the elements of the system, their interrelationships and the associated communications requirements.

Additionally, because of the need for a user centred approach to the deployment of the technology, the design approach and methodology adopted must be capable of emphasising user requirements within the design model. Users themselves fall into a number of categories of which the main groups considered are as follows:

- The elderly who wish to remain within their own homes and to be independent within their own community while receiving the necessary support to enable them to do so.

- Patients wishing to return home as soon as possible following hospitalisation but who still require some degree of monitoring.
- Physically disadvantaged individuals who wish to operate as independently as possible within their home environment and elsewhere.

The technical goal of achieving an integrated system requires consideration of a number of factors including:

- The provision of appropriate, highly reliable, non-invasive and non-obtrusive sensors to provide the necessary monitoring functions.
- The provision of effective and appropriate communications, both within the domestic environment and externally.
- The provision of an appropriate level of intelligence (machine, human or a combination thereof) at each point within the network.
- The integration at the system level of the

Table 1
Systems technologies

Area of Technology	**Comments**
Personal	At the personal level the role of the deployed technology is that of providing all the necessary information about the client and of communicating this onwards within the system. Applications at this level include the monitoring of a persons vital signs, for instance during the recovery period following discharge from hospital, the detection of falls and similar events, the monitoring of therapeutic systems such as drug dispensers and the control of embedded systems such as pacemakers and drug pumps.
Lifestyle	Lifestyle technologies are concerned with the ability of an individual to function effectively within their local environment including the home, their place of work and public areas such as stores. Lifestyle technologies therefore include: • Monitoring systems capable of building up a profile of behaviour and of detecting and responding to anomalies in that profile. • Environmental monitoring covering aspects such as temperature or air quality. • Security systems • Mobility and access support systems including wheelchairs and manipulators. • Dwelling design. • User interface design. This is a particularly important area given that a range of interfaces will be required covering a wide range of user capabilities. • Communications within the local community, for instance through the use of the internet to place people in contact with resources such as libraries or even to talk to their neighbours and to expand the range of social contacts for any individual. • Local intelligence and decision making capabilities to control system operation. • Mobile communications systems such that when a user is within an appropriate environment; home, work or public, they are still connected to the system. Ultimately it is envisaged that the system would become fully mobile, for instance by using satellite technology, enabling all monitoring at the personal level to continue wherever the user might be. • 'Smart' or 'Friendly' home technologies to support the operation of a wide range of devices. This would allow individual items of equipment, cookers, heating systems, therapeutic and physiotherapy equipment to be given individual internet addresses for remote communication and control.
Support	At this level the technologies are concerned with ensuring that the client has access to the support they require as and when they require it. The technologies are therefore concerned with the management of information to ensure that it is presented to the necessary agencies in as up to date form as possible as it is required. This implies some form of intelligent data filter to maintain privacy while still ensuring that the information required to generate the appropriate response is supplied.

individual technologies deployed which may range from alarms to mobility systems such as wheelchairs and 'Smart Home' environments.

Table 1 provides a summary of the technological requirements of the system in relation to the personal, lifestyle and support functions identified.

In addition to the technical aspects associated with meeting the system requirements the ethical and moral questions will need to be addressed including:

- The permitted and acceptable levels of intervention within the domestic environment.
- The empowerment of both the clients and the service suppliers.
- The recognition of the importance and needs of informal carers.
- The effects of demographic change on the consumption and provision of resources and the continuing need to ensure the most effective use of resources and return on investment.

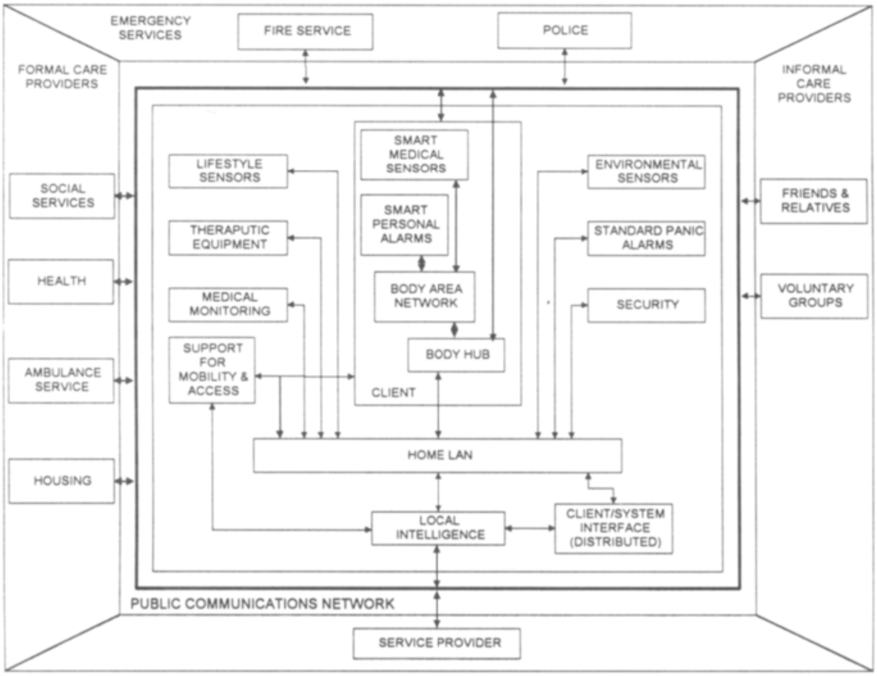

EMERGENCY SERVICES
FIRE SERVICE
POLICE
FORMAL CARE PROVIDERS
INFORMAL CARE PROVIDERS
SOCIAL SERVICES
HEALTH
AMBULANCE SERVICE
HOUSING
FRIENDS & RELATIVES
VOLUNTARY GROUPS
LIFESTYLE SENSORS
THERAPUTIC EQUIPMENT
MEDICAL MONITORING
SUPPORT FOR MOBILITY & ACCESS
SMART MEDICAL SENSORS
SMART PERSONAL ALARMS
BODY AREA NETWORK
BODY HUB
CLIENT
ENVIRONMENTAL SENSORS
STANDARD PANIC ALARMS
SECURITY
HOME LAN
LOCAL INTELLIGENCE
CLIENT/SYSTEM INTERFACE (DISTRIBUTED)
PUBLIC COMMUNICATIONS NETWORK
SERVICE PROVIDER

Figure 1. Major system elements and the associated information flows.

- Social infrastructures and the changes to those infrastructures brought about by technology.

2. SYSTEM ANALYSIS

In order to provide a clear framework for the understanding of system requirements a formal analysis procedure has been adopted. The first stage of this process, which was adapted from a more generalised mechatronics design strategy, was to divide the system into its major constituent elements and associated information flows, as in figure 1. Following this division, a limited form of controlled requirements expression (CORE) analysis was used to formalise ideas and eliminate ambiguities as well as to identify needs and possible conflicts. Figure 2 shows the viewpoint diagram for a typical client resulting from this analysis.

2.1. System structure

From figure 1 it can be seen that the major system elements, are the client, the local environment, the public communications network, the statutory care and support services and the voluntary or informal carers, each of which has its own particular needs and requirements which may be identified by the viewpoint analysis.

Intelligence, is distributed across the system as required and may be considered as 'hard' or machine based and 'soft' or human based as suggested in table 2 and links the following major elements:

- Sensors and monitoring systems

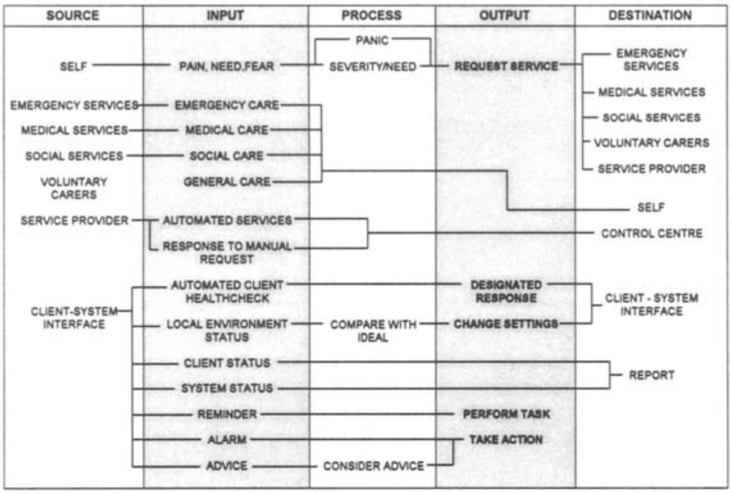

SOURCE
INPUT
PROCESS
OUTPUT
DESTINATION
SELF
PAIN, NEED,FEAR
PANIC
SEVERITY/NEED
REQUEST SERVICE
EMERGENCY SERVICES
MEDICAL SERVICES
SOCIAL SERVICES
VOLUNTARY CARERS
SERVICE PROVIDER
EMERGENCY SERVICES
EMERGENCY CARE
MEDICAL SERVICES
MEDICAL CARE
SOCIAL SERVICES
SOCIAL CARE
VOLUNTARY CARERS
GENERAL CARE
SELF
SERVICE PROVIDER
AUTOMATED SERVICES
RESPONSE TO MANUAL REQUEST
CONTROL CENTRE
CLIENT-SYSTEM INTERFACE
AUTOMATED CLIENT HEALTHCHECK
DESIGNATED RESPONSE
CLIENT - SYSTEM INTERFACE
LOCAL ENVIRONMENT STATUS
COMPARE WITH IDEAL
CHANGE SETTINGS
CLIENT STATUS
SYSTEM STATUS
REPORT
REMINDER
PERFORM TASK
ALARM
TAKE ACTION
ADVICE
CONSIDER ADVICE

Figure 2. Viewpoint diagram for user

Table 2.
Levels of machine-based intelligence

	Level	Sub-System	Information Transfer
HARD	0	Client	*To level 1.* Based on client need.
	1	Sensor	*To level 2.* On request or if an abnormal value or error is detected.
	2	Body Hub	*To level 3.* Usually without further analysis, a validation check may be requested.
	3	LIU	*To level 4.* For detailed analysis if an abnormal or a malfunction is detected. *To level 5.* If an emergency is detected.
SOFT	4	Service Provider	*Information transfer to level 5.* On request or if it is believed that medical or other aid is required.
	5	Emergency Services	-------------

- BodyHUB
- Local intelligence
- Client's healthcare record
- Formal and informal care providers
- Emergency services

The hierarchical nature of this distributed intelligence implies some form of 'information filter' in which information is transferred only as required to limit the total volume of data traffic.

The main role of intelligence within the user environment is that of data management following initial analysis and will only permit onward transfer in response to defined conditions or when agreed data is to be provided on a regular basis, for instance for clinical reasons.

The provision of intelligence within the user environment limits the impact on the public communications network while its distributed nature also has implications for system integrity.

2.2. Viewpoints

Referring to figure 2 it can be seen that the use of an approach based on the viewpoint diagram is effective in forcing the designer into a consideration of the user requirements as part of the overall system structure and supports the identification of key system features. Thus the need for the provision of a 'client-system' interface structure which supports communication from and to the user is readily identified as is the need for direct communication with a range of support and other services in response to defined conditions.

2.3. System development

From figures 1 and 2 it can be seen that the system exists at a number of levels and work is currently in progress towards a demonstrable structure which incorporates a range and variety of sensors from among those set out in table 3. In the first instance consideration will not be given to implementing these sensors in an ideal form but rather to the demonstration and proving of the general system principles and in particular of the

Table 2.
Some physiological and medical sensors

Sensor type	Function
ECG	Pulse rate and variability
Photoplethysmograph	Pulse rate and blood velocity, profile, blood oxygen content
Spirometer	Respiration rate, peak flow, inhale/exhale ratio
Sphygmomanometer	Blood pressure
Thermometer	Basal temperature
Galvanic skin response	Sweating
Colorimeter	Pallor, throat inflammation
Pupilometer	Light response
Accelerometer	Fall and tremor
Polarimeter	Blood glucose level
Stethoscope	Heart and breathing sounds

ability of the system intelligence to interpret and combine information from a variety of sources.

Other work is concentrating on the link between the home environment and the providers such as the operators of control centres linked to community alarm systems. In this area, particular consideration is being given to the ability to rapidly sort and respond to calls of different types requiring different responses, including emergency responses.

3. CASE STUDY

Consider the case of an individual who is being released from hospital following a heart attack. Provided that they are medically stable then it is generally in their best interests to recover at home provided a sufficient level of monitoring appropriate to their condition can be provided. A basic level of provision may be:

- An implanted microsystem to monitor heart rate, blood pressure and blood chemistry.
- An intelligent drug compliance monitor and dispenser which is linked to the hospital and GP data base and which can be adjusted remotely.
- A mobility monitor.
- Internal and external communications systems.

Though this list is not comprehensive, it does give an indication of the level of provision needed and of the associated data collection and analysis requirements. In terms of establishing the user's condition, the following parameters could be transferred to the relevant nodes and devices in the network.

- Bounds for mean and peak values of systolic and diastolic blood pressure.
- Bounds for heart rate.
- Dose rates for prescribed drugs.
- A "ball-park" mobility "score" giving the approximate value of the client's expected mobility back in the domestic environment.

If following their release the client follows the anticipated recovery profile the level of monitoring would be adjusted accordingly until it was removed entirely. However, should a problem be detected then a warning together with instructions would automatically be given to the client and, if appropriate, assistance summoned.

4. FUTURE DEVELOPMENTS

It is increasingly clear that within the next few years there will be a significant increase in the deployment of technologies to support individuals to live independently in their own environment. Many of the technologies involved in this development will be essentially mechatronic in nature and will range from advanced and intelligent wheelchairs to diagnostic and monitoring systems and items of equipment such as intelligent drug dispensers capable of monitoring dosage and use and which are able to be adjusted remotely. For instance, by a patient's GP following the detection and reporting of changes in the patient's condition by the local monitoring systems. Other areas of development include advanced sensors for blood gas analysis and the use of nanotechnologies to develop a range of implants for physiological monitoring and the control of direct drug release.

Many, if not all, of these systems and sub-systems are mechatronic in nature and it is therefore essential that an integrated system framework is created which will fully support their application and use.

4. CONCLUSIONS

A systems approach has been used to identify the main constituent elements of an integrated and intelligent telemedicine system integrating a range of intelligent and essentially mechatronic technologies. The analysis has been based on the use of a modification of the CORE methodology and structured to position the client or user as the centre of the design process. The viewpoints of the other major system elements, both human and technology based, have also been considered and any conflicts identified.

The introduction of hierarchic and distributed intelligence at all levels throughout the system provides a powerful information management feature which improves the ability of the system to provide the required levels of support and which allows for the integration of new technologies as they become available.

The systems level approach adopted is therefore considered as providing a firm base for further development.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Engineering and Physical Sciences Research Council in carrying out this work.

REFERENCES

1. Congress, Office of Technology Assessment, 1995, “Bringing Health Care Online: The Role of Information Technologies” OTA-ITC-624 Washington D.C., U.S. Government Printing Office, September.
2. K. Doughty, K. Cameron, P. Garner, 1996, "Three generations of telecare of the elderly" *Journal of Telemedicine and Telecare* Vol. 2 No. 2 pp 71-80.
3. B. Celler, T. Hesketh, W. Earnshaw, E. Ilsar, 1994, “An instrumentation system for the remote monitoring of changes in functional health status of the elderly at home” *Proceedings of the 16th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, Baltimore, USA, pp. 908-909, Nov. 3rd - 6th,.
4. P. Sharkey, 1995, “Introduction to Community Care” *Collins Educational,* London, pp. 16-32.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Redundant Manipulator as a Tool for Disabled People in an Unstructured Environment

Magnus Olsson, Per Hedenborn, Ulf Lorentzon and Gunnar Bolmsjö

Department of Production and Materials Engineering, Lund University
P.O. Box 118, 221 00 Lund, Sweden

Abstract. This paper is a presentation of a lightweight service robot. The manipulator is basically designed for use with an electric wheelchair and will compensate for upper limb immovability but the modularised structure makes it possible to adapt the arm for new application areas. Furthermore are some aspects of the development phase presented. Techniques to integrate the development of both mechanical and software modules for the robot are utilised. A simulation tool is used throughout the work both for; virtual prototype design, virtual scenarios to verify usefulness, dynamic analyse, kinematics studies, integrating and testing of the physical prototype through telerobotics.

1 Introduction

A lightweight manipulator basically designed to be used in the field of service robotics is presented in this paper. The first prototype is adapted for disabled people and is used with an electric wheelchair as a moveable base. The manipulator will compensate for upper limb immovability of the user. The prototype of the manipulator has 8 degrees of freedom.

The main reasons for the redundant configuration is the possibility to avoid obstacles in an unstructured environment and to decrease acceleration in each joint to minimise sudden unexpected movements of the arm in singular areas. The redundant configuration of the manipulator makes the inverse kinematics crucial to fully utilise the additional degrees of freedom to avoid obstacles in an unstructured environment. The development, implementation and verification process of the inverse kinematics module is described.

The manipulator is designed in different modules. This will enable tailor-made solutions for different users and applications. By making self-contained links with distributed control and communication between the modules through a field-bus, the modularity of the robot is further enhanced.

The safety considerations are especially important for this robot where the man and machine is working closely together. The possibility to reverse the joints by hand-power and intelligent software for control gives a safe solution altogether.

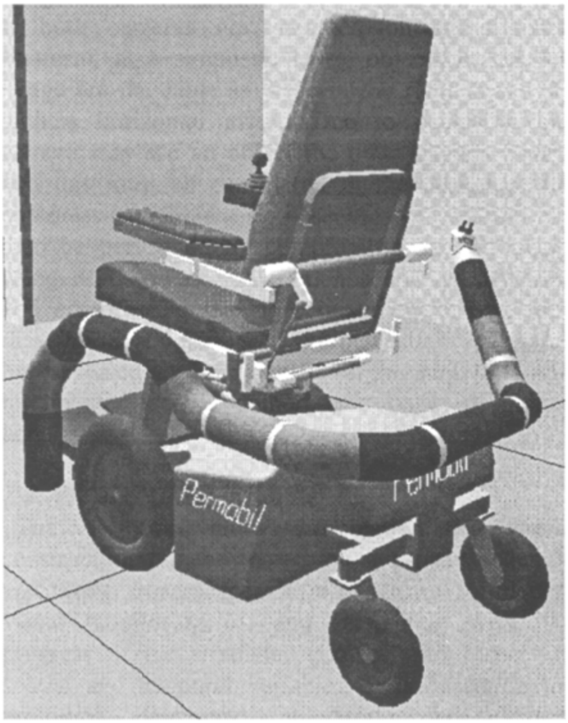


Figure 1. A snapshot taken in the simulation software of the robot arm mounted on an electric wheelchair.

The development of the robot has been made with a concurrent engineering philosophy. Several prototypes have been tested in a robot simulation system TELEGRIP[1] and verified before a mechanical prototype even existed. In figure 1 a snapshot of the robot arm from the simulation software is presented.

Controller software have been developed and tested in the simulation system. The usability of different robot configurations was tested in virtual scenarios in typical applications, e.g. office environment [1]. TELEGRIP has also been used as a top-level controller in the early test of the physical prototype.

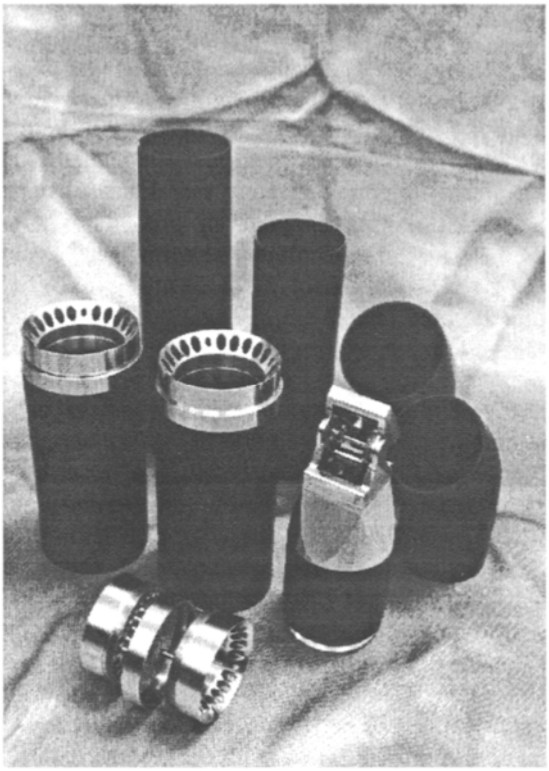


Figure 2. Different components of the robot arm made of lightweight materials like aluminium alloys and carbon fibre.

2 Design of the manipulator

This manipulator is aimed for the field of service robotics. An application that is believed to have a good market potential in the near future. These service robots will cover a new area of application and have different criteria's than standard industrial robots.

- Payload in lower range (typically 1 kg) and the payload/weight ratio have to be better than industrial robots.
- Low repeatability demands since the user close the control loop.
- Flexible configuration and a wide range of application areas.
- Lower life duty cycle, velocity and accelerations compared to industrial robots.
- Must have low production cost at high volumes.

The chosen solution have some design features that not are so commonly used on industrial robots: carbon fibre composite links; self contained joints that include motors, reducers, bearings and control circuits; field-bus communication that handles positioning and feedback information. Some of the structure components are shown in figure 2. Since the manipulator is intended to be mounted on an electric wheelchair it is very important that it is able to fold up in a position within the chairs outer limits.

The field-bus solution for communication between the different modules enable a minimal number of wires through the arm that otherwise would complicate the assemble and dis-assemble of the arm. Another feature is the simple mechanical structure of the joints. Together with the servo a planetary reducer is integrated and no other mechanical transmission is needed. Figure 3 shows a cross-section of a typical joint in the robot arm.

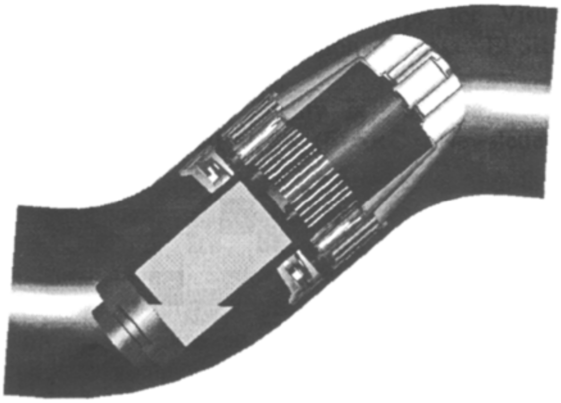


Figure 3. A cross-section of joint 5 in the robot arm prototype. Servo, encoder and reducer are integrated.

---

[1] From Deneb Robotics Inc.

After several virtual prototypes in the simulation system TELEGRIP, a fairly good idea of how the manipulator should look like was formed and made it possible to start the design on component level rather early. In [2] some of the techniques and benefits of the parallel work is presented.

3 Distributed servo control

Both control and servo is distributed in the arm. Each module of the arm contains one to three brushless DC servos, servo amplifiers, PID-control circuits and a micro-controller for field-bus communication.

The electronics are custom-made for the different kinds of robot-modules. The choice of the electronics is based on the available space in the modules and the power demands of the servos.

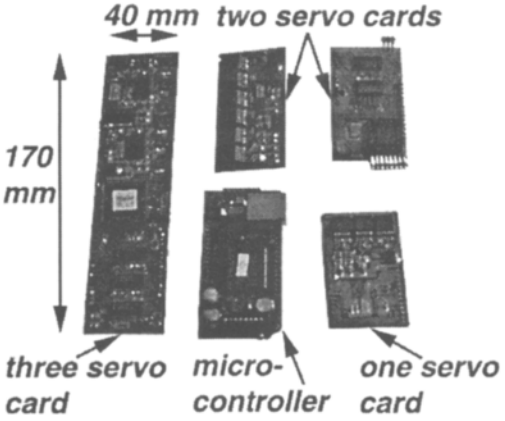


Figure 4. The electronics board developed for the robot arm.

The micro-controller receives information via the field-bus for each servo about goal-positions, velocity, acceleration and new PID parameter settings. The PID circuit calculates the signal to the servo based on information about the current position of the motor obtained from the optical encoders. This signal is commutated and further amplified. All functionality is on a single board.

The micro-controller can also check status for each servo and transfer it back to the supervisor controller. Some safety issues can also be imposed to the micro-controller. For instance if the difference between the desired and the real position is increasing, which could be a sign on a movement against a obstacle, the micro-controller could take own decisions and shut off the motor or increase the regulator parameters.

The electronics cards used in the arm is presented in figure 4.

4 Development of inverse kinematics routine and verification in simulation system

4.1 The use of TELEGRIP

The simulation system has been used throughout the development cycle of both mechanics and software. Examples are dynamic analyse used for dimensioning the servos and virtual scenarios used for deciding the configuration and veryfing the usability. For this redundant eight joint prototype arm the inverse kinematics routine is especially important. TELEGRIP has turned out to be a valuable tool for the development of this routine.

Writing an inverse kinematics routine for a new robot configuration is often a tedious task. There are several possible pitfalls in the work to match the routine view of the robot with the physical. The validation of the routine without a visual simulation model is also intricate.

By using TELEGRIP and create a virtual model of the new manipulator the inverse kinematics routine can easily be tested. By implementing the routine in a shared library, the routine can then be called in run-time to calculate the joint-values for a requested position and orientation of the robot tool. The creation of the trajectory is still taken care of by TELEGRIP.

4.2 The inverse kinematics routine for the presented robot arm

For the presented redundant robot now obvious way exist to calculate the inverse kinematics analytically. A numerical and iterative solution is used instead, based on the algorithm presented in [3]. This method is adapted to the case of the redundant robot presented in [4].

The method is based on a numerical calculation of a residual vector describing the difference between the actual pose (position and orientation) and the goal pose of the manipulator. The residual vector is denoted as $\boldsymbol{r}$. The actual joint values of the robot are denoted as the vector $\boldsymbol{q}$ and in the goal pose the unknown vector $\boldsymbol{q}^*$. The solution is obtained when the actual tool transform $\boldsymbol{T}^a$ is equal to the goal tool transform $\boldsymbol{T}^t$. This is only the fact when $\boldsymbol{q}=\boldsymbol{q}^*$ and the residual vector becomes a zero vector, i.e. $\boldsymbol{r}(\boldsymbol{q}^*)=\boldsymbol{0}$. The algorithm to search for the solution can be written as;

$$\boldsymbol{q}^{(k+1)}=\boldsymbol{q}^{(k)}+\delta^{(k)} \qquad (1)$$

The term $\delta^{(k)}$ can be written as;

$$\delta^{(k)}=-\alpha \boldsymbol{J}^{+}\boldsymbol{r}(\boldsymbol{q}^{(k)})-(\boldsymbol{I}-\boldsymbol{J}^{+}\boldsymbol{J})\nabla \boldsymbol{H}(\boldsymbol{q}^{(k)}) \qquad (2)$$

J is the Jacobian matrix of the robot and has the size m×n , where m is the number of kinematics equations (constraints) and n is the manipulators degree of freedom. It is basically a type of Steepest-Descent algorithm. The $(\boldsymbol{I}-\boldsymbol{J}^{+}\boldsymbol{J})$ matrix is the null space of the Jacobian matrix. The appearance of the $\mathbf{H}(\mathbf{q})$ matrix do not effect the motion of the tool and can be used to add extra constraints as;

- Avoid mechanical joint-limits.
- Avoid collisions with obstacles by defining forbidden space.
- Avoid singularity by maximising manipulability.
- Minimise joint torque and accelerations.
- Minimise power consumption.

Extra criteria like avoiding mechanical joint limits is suggested in [5] to be expressed as a row in the $\mathbf{H}(\mathbf{q})$ matrix in the following way;

$$\mathbf{H}(\mathbf{q})=\sum_{i=1}^{n}\frac{1}{4}\frac{(q_{i,\max}-q_{i,\min})^2}{(q_{i,\max}-q_i)(q_i-q_{i,\min})} \qquad (3)$$

In figure 5 an example plot of a constraint function is presented for a case where the limits for a specific joint is minus 60 degrees and plus 80 degrees. The constraint function increases drastically close to the specified limits. The algorithm is trying to minimise these functions and will avoid the limit.

The use of TELEGRIP has simplified fast and accurate validation of the motion accomplished with different constraints and combination of constraints. It is especially important to study combinations of constraints since these can cause singularities introduced by the algorithm itself.

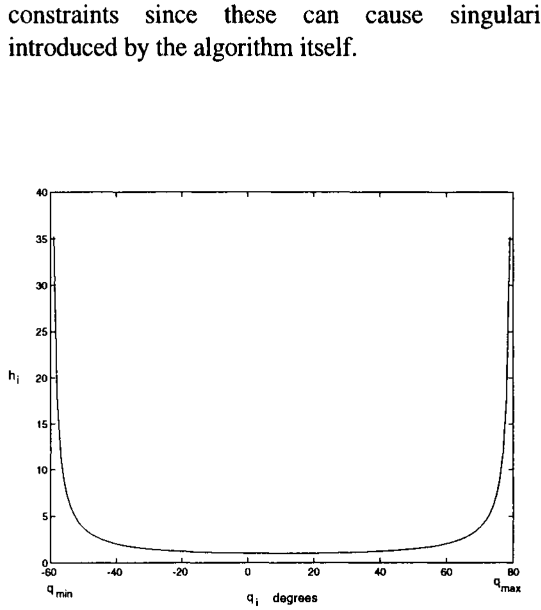


Figure 5. The constraint function for avoiding mechanical joint limits.

To be able to efficiently develop new algorithms for different robot configurations, the code generation for the calculation of residual vector and Jacobian matrix is automated in MapleV[2]. The mathematical models describing the robot structure are often very lengthy. Deriving such expressions by hand are very time consuming and error-prone. The approach to use MapleV to symbolically describe the robot and generate C-code can be used in several other areas of robotics as dynamics, trajectory planning and control. In [6] a survey of the research in this area is presented.

5 Initial testing of physical robot prototype through tele-robotics

The integration process of the mechanical prototype and the controller software can be a very time consuming task. This process should start as early as possible to discover defects in the beginning of the development process. By using the simulation system as the primary controller, and transferring low-level manipulator information (joint values and velocities) to the target system for the control process, the pre-defined modules in the simulation system can be used to supply the functionality that still is missing in the target system.

---

[2] Waterloo Maple Inc.

TELEGRIP is especially made for these kinds of systems. LLTI (low-level tele-robotics interface) is used to transfer external information to the simulation system and change the proceeding simulation by external data. Furthermore LLTI could be used to extract data from the simulation system and propagate it to external systems. The maximum frequency for this data exchange is 30 Hz. The LLTI module is a procedure that automatically is executed 30 times per second. There is complete flexibility to add functions in this code that is provided by the IRIX OS. Both TCP/IP and serial communication from the LLTI module have been tested.

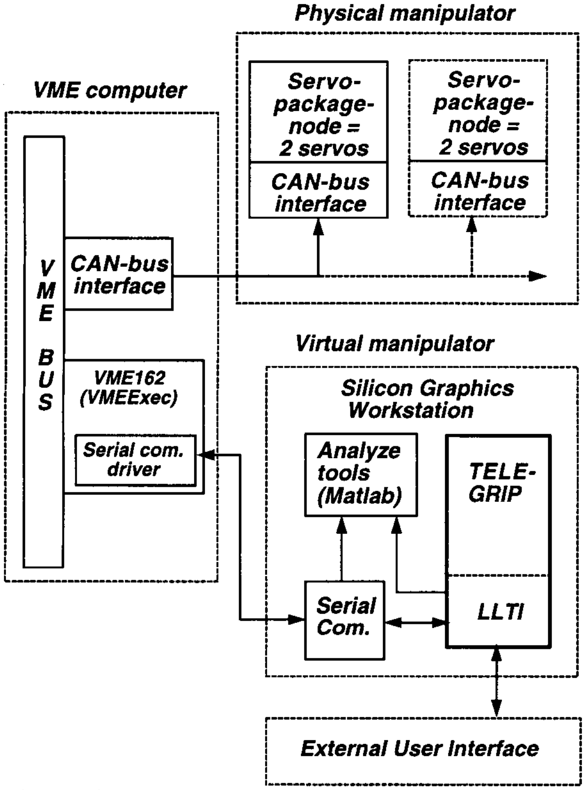


Figure 6. The structure of the tele-robotic system used for validation of the mechanical robot prototype.

Examples of how TELEGRIP can be used for real-time control are presented in [7]. This paper also discusses the use of sensors for continuously updating the virtual world model. The simulation model can then be used for decision making, re-planning and adaptation of the tasks to the current state of the world.

In the application of manipulator validation the only feedback from the physical manipulator is the resolver data reflecting the state of the individual joints of the manipulator. This information can be used to check the response for the applied control algorithm. It is also possible to have external analysing tools running on the workstation.

In figure 6, the experimental system for the wheelchair manipulator is presented. Four CAN-bus interfaces are located in the arm, and each interface is controlling one to three servos. In the VME-system, the main controller, a fifth CAN-bus interface is located and is working as a master, distributing information to the different servos and receiving information of the actual motion. The technique is highly modularised and new configurations are easily provided. When the mechanical configuration is changed, the software modules have to reflect the changes. This results in high requirements on the adaptability of the software modules.

The test of new user interfaces for the control of the robot arm is also possible through the LLTI interface in TELEGRIP. The user interface can be tested solely on a virtual robot model in TELEGRIP or on the virtual and the physical robot using the telerobotics interface.

6 Conclusions

In the development of robots and especially in the field of service robotics, parallel work with both mechanical prototyping and controller software design is important. The time gain could be considerable if the simulation tools are reliable. In the presented case study different aspects of the development process is considered.

The tools are useful no matter how the manipulator configuration is chosen. The simulation tools should be used as early as possible. Already in the outline of the first virtual prototype the tools could be of powerful help. The tools are then used throughout the process with the final integration and verification of the physical manipulator as a final step.

The inverse kinematics algorithms, the process of converting the robot task described in the Cartesian space to joint space, are investigated. For the redundant manipulator presented the need of extensive testing is especially emphasised.

During the integration phase, the simulation tool can be used as well. Now as a master for the manipulator, feeding the target system with low-level joint state information. In this way, the time from idea to a working manipulator is reduced.

7 Acknowledgements

The work presented in this paper has been funded by AMS (The Swedish National Labour Market Board), RFV (National Social Insurance Board) and NUTEK (Swedish National Board for Industrial and Technical Development). Part of the work has also been funded by the European Commission under Contract ERB FMRXCT 960070 (MOBINET).

References

1. G. Bolmsjö, et. al. Simulation Based Design of a Wheelchair Mounted Robot for Disabled People, Deeneb user conference, Troy, MI, USA, Oct 7-11, 1996.
2. U. Lorentzon and M. Olsson. Eight Axis Lightweight Service Manipulator: Virtual Prototyping of Mechanical Structure and Controller Software, World Congress Manufacturing Technology Towards 2000, Cairns, Australia, Sep 15-17,1997.
3. A. A. Goldenberg, et. al. A Complete Generalized Solution to the Inverse Kinematics of Robots, IEEE Transaction on Robotics and Automation, 1(1) 14-20, 1985.
4. M. Olsson, et.al. Simulation based Validation of Controller Software in Robotics, World Congress Manufacturing Technology Towards 2000, Cairns, Australia, Sep 15-17,1997.
5. T. F. Chan and R. V. Dubey. A Weighted Least-Norm Solution Based Scheme for Avoiding Joint Limits for Redundant Joint Manipulators, IEEE Transaction on Robotics and Automation 2(11) 286-292, 1995.
6. N. Rentia and N Vira, Why Symbolic computation in Robotics?, Computers in Industry, 17(1) 49-62, 1991.
7. P. Hedenborn, et.al. Sensor controlled robotics arc welding integrated with simulation software, Mechatronics 98', Skövde, Sweden, Sep 9-11, 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

A HELMET MOUNTED DISPLAY SYSTEM WITH ACTIVE GAZE CONTROL FOR VISUAL TELEPRESENCE

Julian P Brooker BSc
Paul M Sharkey BSc, MA, PhD, CEng, MIEE, SMIEEE
Interactive Systems Research Group
Department of Cybernetics, University of Reading, Whiteknights, Reading, RG6 6AY, UK

ABSTRACT

Visual Telepresence systems which utilise virtual reality style helmet mounted displays have a number of inconvenient disadvantages. The geometry of the camera positions and of the displays is fixed and is most suitable only for viewing elements of a scene at a particular distance. To address such limitations a prototype system where the geometry of the displays and cameras is dynamically controlled by the eye movement of the operator has been developed. This paper explores why it is necessary to actively adjust the display system as well as the cameras and justifies the use of mechanical adjustment of the displays as an alternative to adjustment by electronic or image processing methods. The electronic and mechanical design is described including optical arrangements and control algorithms. The performance and accuracy of the system is assessed with respect to eye movement.

1. INTRODUCTION

In a visual telepresence system which utilises a standard helmet mounted display slaved to a robotic head the operator's ability to gaze around a scene without the use of head movement is severely limited. The operator will also experience difficulty when attempting to fuse the image of an object which is at a non-ideal distance. With fixed accommodation, the cross-coupling of the brain's control systems for vergence and accommodation will also lead to stimuli-conflict. An additional reason for the poor quality of telepresence systems is due to the low resolution currently obtainable in small size LCD displays. A trade off must be made between a poor resolution or a narrow width of viewing field. Within virtual reality systems, helmet mounted displays are designed to optimise the sense of immersion felt by the operator and to this end a large width of viewing field is employed. Within telepresence applications however, immersion may be a secondary issue to the ability of the operator to resolve fine detail.

The usual camera configuration used within a visual telepresence system is orient both cameras in a parallel direction resulting in a 100% binocular overlap at infinity. A good sense of natural perception of both depth and scale will be experienced by the human observer provided that the following conditions are met [1]:

- The Inter-Camera Distance (ICD) and Inter-Display Distance (IDD) match the observers Inter-Pupil Distance (IPD).
- The Field of View (FOV) of the camera/lens configuration is the same as the FOV of the display/lens configuration.

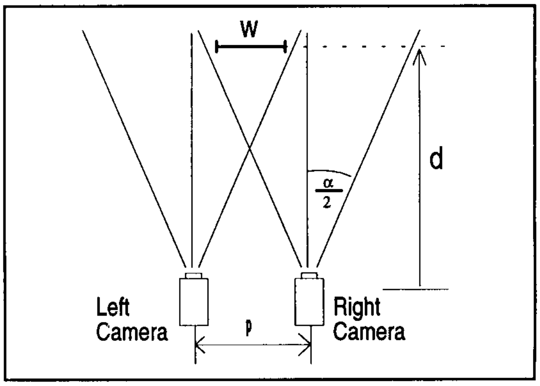


Figure 1. *Relationship between FOV and overlap given fixed ICD*

In potential applications of telepresence, a high quality binocular image of elements in the near viewing field is more important than a wide-angled view of the distant viewing field. The logical approach to solving this problem is to reduce the FOV of the cameras and displays. As the FOV is reduced, the percentage of overlap occuring on an object in the near viewing field also decreases. This is illustrated in figure 1, the minimum distance d at which the target is entirely in viewed is:

$$d = \frac{p + w}{2 \tan \alpha / 2} \tag{1}$$

where

w = width of the target
p = IPD of the operator
α = horizontal angle of view

A means of optimising binocular overlap at close range whilst using a small FOV would be to verge the cameras inwards until they centre on the region of interest. A problem with this approach is that the relationship between the vergence angle of the operator's eyes required to fuse the image and the distance of the object would break down, as illustrated in figure 2.

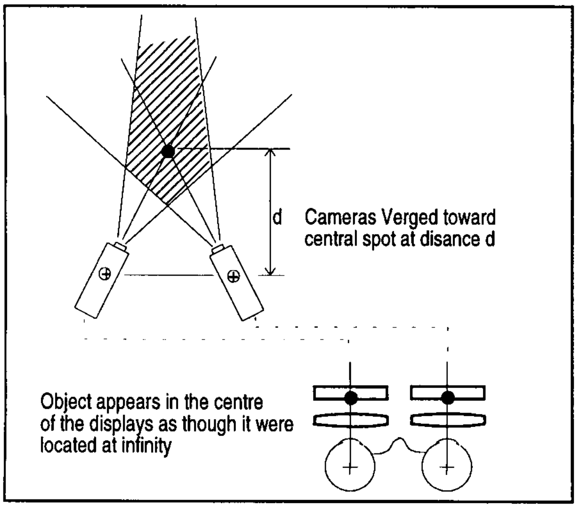

d
Cameras Verged toward central spot at disance d
Object appears in the centre of the displays as though it were located at infinity

Figure 2. *The breakdown between vergence and distance with non-parallel cameras.*

2. DYNAMIC VERGENCE BY USE OF EYE TRACKING

A potential solution to the optimisation problem would involve tracking the operators eye movements, as a result of which the camera geometry could be optimised for the distance and azimuth of the area of interest. A critical item in such a system is the display [2]. Figure 3(a) shows the eye stabilised on a target. In (b) the target moves and accordingly the image moves on the display. In (c) the eye moves to fixate the target. Via the eye tracking system the camera changes position(d). In diagram (e) the eye is pursuing the target again as its display position has altered.

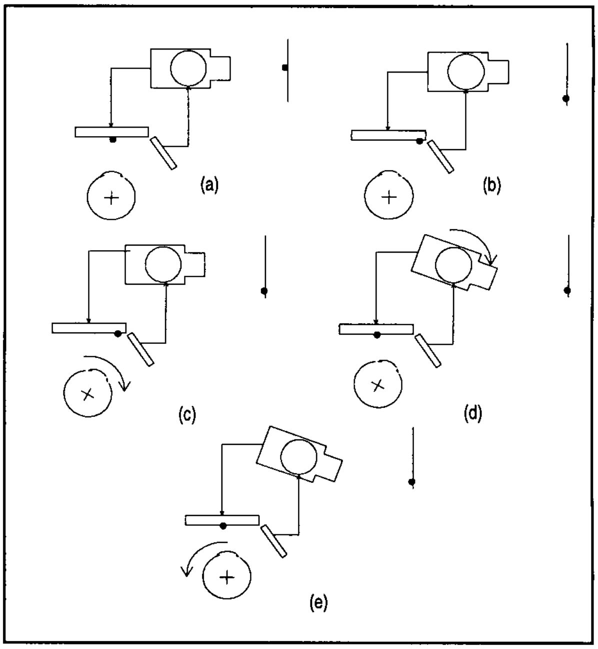


Figure 3. *The need to move the display.*

For the system to function in such a way that a target will have a stable fixation point, the displayed image must move by the same angle as, and synchronously with the camera via either electronic or mechanical displacement. Neither of these approaches would be particularly simple to achieve. Initially, the approach of using electronic image translation seemed a potentially better solution than a mechanical one, the reasons being:

- No mechanical moving parts would be required.
- Horizontal image deflection could be achieved with minimal additional image processing

hardware, e.g. by adding an offset to the rastering circuitry of an LCD display driver.

- The dynamic performance of an electrical translation system is likely to be superior to a mechanical system.

There is a major disadvantage to using an electronic method to displace the image, however, and that is the low resolution available in LCD displays. Using a method of active area translation would considerably decrease the image resolution. This would outweigh any advantage to be achieved from using active camera geometry.

3. MECHANICAL DISPLAY POSITIONING

A more suitable method than electronic displacement of the image would be to make each eye display moveable. If the displays were to be mobile however, it would have to be in such a way that they were always directly facing the pupil i.e. each display would have to rotate about the eye's centre. If the display was simply moved laterally in front of the eye then the eye would be viewing the optics at an angle and optical distortion from the lenses would be greatly increased, the displayed images require stretching via image processing to give a correct perspective. Figure 4 shows the type of motion required. The effect of the eye moving to an off centre point of interest would be to cause both the camera and the display to move by the same number of degrees. This would mean that the point of interest would remain in the same angular position in front of the eye.

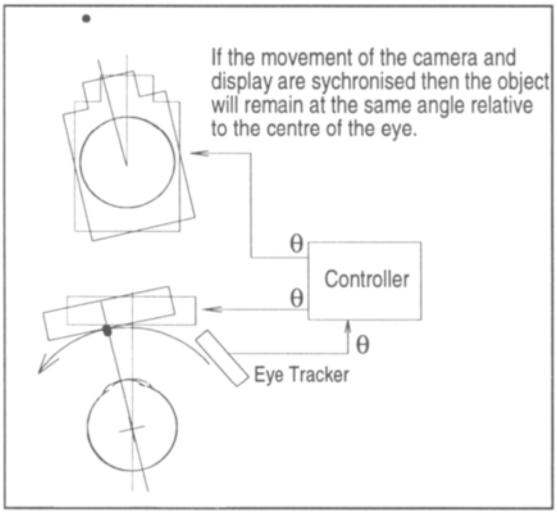

If the movement of the camera and display are sychronised then the object will remain at the same angle relative to the centre of the eye.
θ
θ
Controller
θ
Eye Tracker

Figure 4. *The required display motion.*

Potential problems with the design of such a system are that if the eyes are converged to view a region in the near display field the displays would collide and also that the desired centre of rotation for the display path is the centre of the eyeball, this would require locating a rotational pivot above the operator's head.

A former approach to this problem was to use a pantagraph assembly to produce a displaced centre of rotation[3]. A better solution has been produced by the use of mirrors to create a new centre of rotation for the displays in front of the operator's head. The display system is illustrated in figure 5.

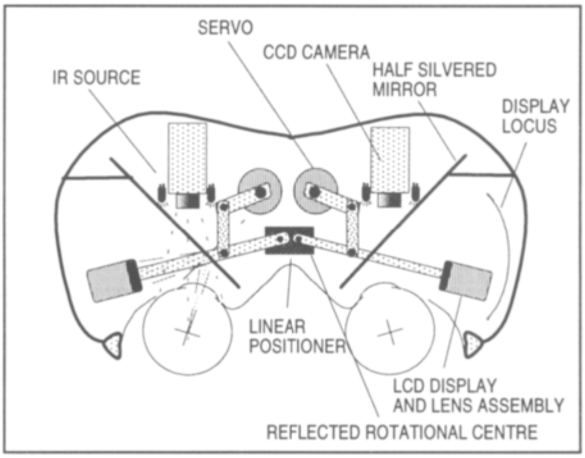


Figure 5. *Prototype Display System*

To monitor the vergence angle of the eye, infra-red tracking is employed. The eye is illuminated with a broad beam of invisible infra-red (IR) light. Since the outer eye is a poor reflector of IR light and the retina is a good reflector, a spot is formed on the CCD array. Image processing is used to identify the spot and locate its centre, from this information the azimuth of the eye is calculated. The positioning system consists of a dc-servo driven arm upon which the eye display is mounted. The control system attempts to keep the orientation of the display perpendicular to the eye's line of sight reflected in the mirror and at a fixed distance from the pupil. The controller also simultaneously adjusts the vergence angle of the camera via servos at the remote site to match the viewing angles of the eyes. The simultaneous adjustment of the cameras ensures that the viewed image does not appear to translate in front of the operator's eyes.

The resultant system ensures that the angle of the cameras is triangulated to the point in the horizontal plane at which the operator is viewing. The angles in which the operator's eyes look correspond to those produced when viewing the scene first hand, ensuring that the image may be easily fused and avoiding stimuli conflict. Since the displays are centred on the horizontal line of sight, optical distortion from the lenses in the camera and display systems will be minimised.

4. CONTROL SYSTEM

4.1 HARDWARE CONFIGURATION

A single-eye prototype display and camera has been developed in order to test the control system. The prototype display has an angle of view of 30° which can be moved over a range of 57°. The eye-tracking system has an accuracy of 30 arc minutes, a detection range of $^+/-$ 30° and an update rate of 50 Hz. The output from the eye tracking system is fed to a microcontroller via an RS-232 link (figure 6). The microcontroller commands two servo controllers which are based on LM629 PID controller ICs. The two servos are of identical type and the display and cameras have similar inertial moment. The maximum acceleration achievable is approximately $10000^\circ\ s^{-2}$.

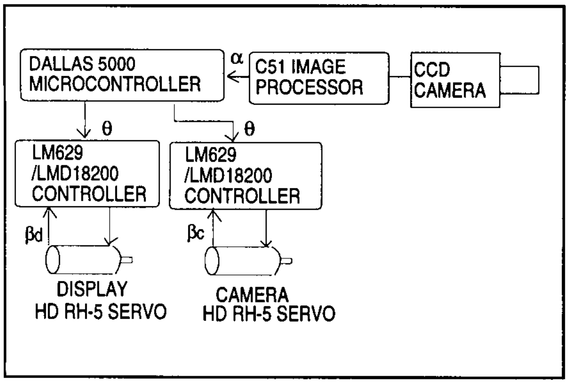


Figure 6. *Control System for one Camera/Display.*

4.2 EYE MOVEMENT

Due to the complexity and different modes of eye movement, it is necessary to carry out some processing of the data from the eye tracker to optimise the performance. There are four main types of eye movement, three of which are conjugate, and one disjunctive[4]. The varieties of conjugate movement are:

- Pursuit - during visual pursuit of a moving target the head is normally stationary whilst the eyes attempt to maintain fixation on the moving target. If the eyes lose track of the target, quick phase saccadic eye movements will occur to reposition the eyes.
- Opto-Kinetic - looking at constant motion of large visual fields, such as looking out of the window of a train. This movement consists of repeated periods of slow phase movement followed by a quick phase flyback.
- Vestibulo-Ocular - this motion occurs when the head rotates but the eyes remain stabilised on a target.

In conjugate movement, slow phase movement ranges from 1 to $30^\circ\ s^{-1}$ attempt to stabilise the image of the moving target on the retina. If slow phase fails then quick phase saccadic movements reposition the eye. The magnitude of these movements is in the range of 1 to 40°. Acceleration can be as high as $40000^\circ\ s^{-1}$ and the peak velocity can reach $500^\circ\ s^{-1}$. The duration of a saccade is 30-100 ms.

Disjunctive movement (vergence) is where the eyes move in opposite directions. This will occur when the eyes alter their fixation from one target to another target at a different distance. Disjunctive eye movement is slower and smoother than conjugate movement with a maximum range of about 15° (target at infinity to target at 10 cm) and a velocity of less than $10\ ^\circ\ s^{-1}$.

There are a number of very small movements made by the eyes when they would otherwise be stationary, these are:

- Flicks (micro-saccades) - saccade like darts of movement less than 1° in magnitude.
- Drifts - random direction, small and slow movements, $< 0.1\ ^\circ\ s^{-1}$.
- Tremor - a small high frequency oscillation < 30 arc seconds, frequency 30-150 Hz.

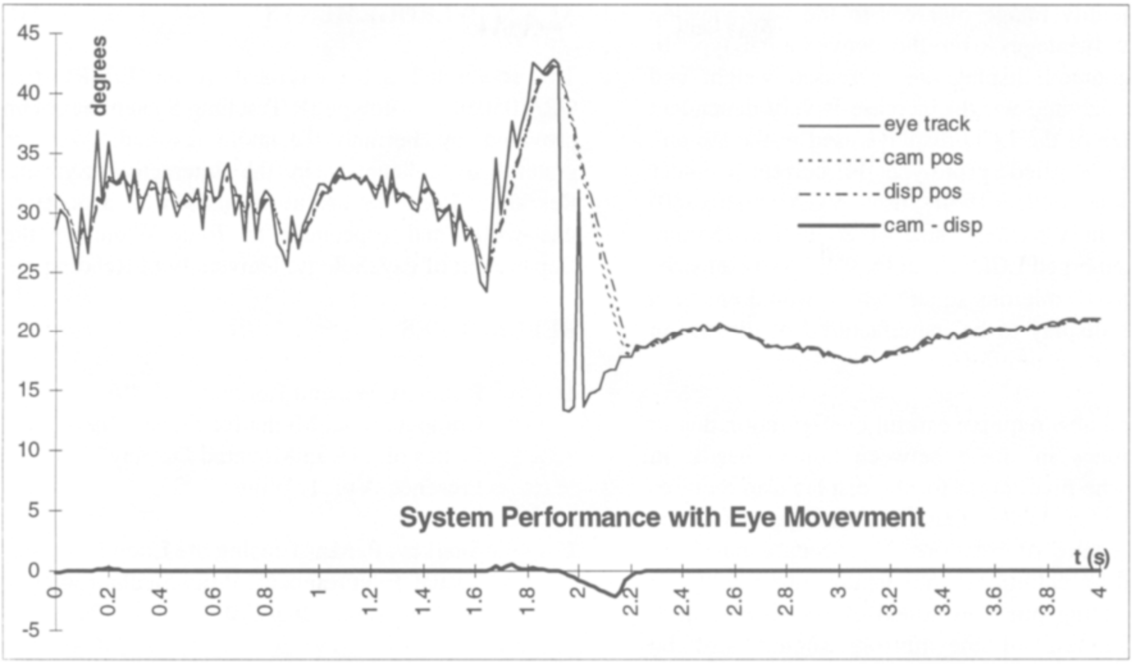


Figure 7. *Response of System to human eye movement.*

4.3 CONTROL ALGORITHM

The objective of the control algorithm is to keep the display position centred on the eye whilst synchronising its movement to that of the camera. The display and camera are driven by the same type of servo which consists of a DC motor fitted with an 80:1 low backlash Harmonic Drive reduction gearbox and a 360 line optical encoder. Since the inertial moment of the camera and the display system are very similar, the performance of both axes is near identical and the difference error between them is small. An iterative process was used to tune the parameters of the LM629 PID controller to give a critically damped response on the camera axis. The same control parameters were applied to the display axis although more precise tuning may reduce the positional error under high velocity movement. It also desirable to limit unnecessary movement of the display which could unsettle a stabilised eye position. The effects of minor movement from the eye are reduced by application of filtering to the data from the eye tracker.

5. TESTING OF THE PROTOTYPE SYSTEM

Figure 7 shows data from the system with a human operator viewing a scene containing a number of moving objects. Over the section shown the eye tracks rapidly about the displayed image and finally fixates on a moving target. The result of subtracting the display position from the camera position shows that the overall disturbance in the position of the image presented to the eye is minimal except in the case of a large saccade. After a large saccade to a new position it will take some time for the display/camera system to catch up due to the differences in acceleration, the eye however, takes some time to settle and re-stabilise the image and the transient error in the image position may not be perceptible.

6. ERGONOMICS AND USABILITY

The prototype system allowed an approximate doubling of the spatial resolution for the equivalent maximum horizontal angle of view in a fixed display helmet and allowed the eye to scan the scene in a horizontal manner. Such a system would be particularly suited to telepresence tasks which require a good quality image of areas in the near viewing field. Disadvantages over the conventional type of helmet mounted display are increased weight and size. The size and weight increase is very dependant on the size of the LCD displays used in the system. A full two sided prototype is currently under construction which will use VGA (640x480) resolution in a package size of 20 mm x 15 mm. Recently emerged LCD technology allows pixel sizes as small as 10 microns square which would enable a 1024x768 display to be manufactured of dimension 12.5 mm x 12.5 mm [5][6].

The system also requires careful configuration due to the difference in shape between human heads, in particular the pivot point for the display arm requires adjustment in order that it coincides with the reflected centre of rotation. A procedure has been developed for calibrating the display correctly by use of two sighting pins positioned on the display arm. The positioning of the mirrors should also be adjustable to maximise the viewing range for a specific individual.

7. CONCLUSION / FURTHER WORK

An advantage of this system is that the horizontal width of the viewing field becomes less important since the operator can use gaze to direct the displays to be centred upon the area of interest. The horizontal viewing field may therefore be reduced to yield a higher fidelity display of the area under view whilst supporting the natural tendency of the eyes to scan across scenes horizontally. A further advantage is that it is possible to monitor the activity of the operator's eyes in viewing of the remote scene and provide the information such as a readout of the distance of an object at which the operator is currently looking.

The dynamic camera/display system offers a promising method of extending the capabilities of helmet mounted display systems. A full usability assessment is due to be carried out when the full working prototype is completed.

ACKNOWLEDGEMENTS

This research has been funded by the EPSRC (ref GR/L05054). Prototype IR Tracking System has been provided by Ferranti Technologies Ltd. We are grateful to colleagues in the Interactive Systems Research Group for useful discussion on aspects of this work, and especially to John Wann of the Department of Psychology, University of Reading.

REFERENCES

1 Robinett, W., and Rolland, J.P., "A Computational Model for Stereo Display Optics of a Head Mounted Display", Presence, Vol. 1, Winter 1992.

2 Sharkey, P. M., "Closing the Loop for Visual Telepresence", Proc. Mechatronics 96, Vol. 1, p.p. 285-290.

3 Sharkey, P. M., Murray, D. W., McLauchlan, P. F., Brooker J P., "Hardware Development of The Yorick Series of Active Vision Systems", to appear, Microprocessors and Microsystems Journal, 1998.

4 Carpenter, R.H.S., "Movements of The Eyes, 2nd Edition", Pion, London 1988.

5 Electronic Engineering Times, "LCD on Silicon Displays Turn Heads at SID Confab", January 20 1996.

6 Handschy, M., and Locke, C., "High Performance Virtual Displays", Advanced Imaging Magazine, January 1997.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Integrated Telepresence, Virtual Reality, and Mobile Communications
A Project Report

G. Mair[a], N. Clark[a], R. Fryer[b], R. Hardiman[b], D. McGregor[b], A. Retik[c], N. Retik[c], K.Revie[a]

[a]Transparent Telepresence Research Group, Department of Design, Manufacture and Engineering Management (DMEM), University of Strathclyde, Glasgow G1 1XJ, Scotland, UK

[b]Department of Computer Science, University of Strathclyde, Glasgow G1 1XJ, Scotland, UK

[c]Department of Civil Engineering, University of Strathclyde, Glasgow G1 1XJ, Scotland, UK

This paper describes an interdisciplinary two year project to research and develop an integrated immersive telepresence, virtual reality, and mobile communications system. The system incorporates subsystems such as mechatronic stereo and mono camera platforms, a remotely operated land vehicle, a cellular mobile phone system for video and control communications, home site equipment such as a head mounted display, and virtual reality software. As an application example the system is deployed for the planning, monitoring, and recording of actual against planned progress on construction sites. The planned progress is provided via the virtual reality software and this is compared with real world progress as displayed through the telepresence system. The paper explains the project objectives, progress to date including the system specification, the technology developed, the system integration, the results of initial trials, and future plans.

1. INTRODUCTION

In any major engineering project those involved wish to be able to monitor actual against planned progress, check that work is being carried out to specification, and maintain a record of what work has been done and when and where it was carried out. It is also the case that many of the interested parties may not have easy access to the project site due to either lack of time, or difficulty in accessing the site because of distance or hazardous conditions. This project attempts to satisfy the above requirements and allow all of them to be met while the parties are remote from the site.

1.1 Objectives

It is intended to produce a telepresence system that embodies immersive and non-immersive configurations. It is also planned to include the capability to allow real world images from a remote site to be viewed at a home site combined with information and simulated images.This combination of real and virtual world data will have the potential to allow major engineering projects to be monitored and inspected from anywhere in the world. Initial virtual models have been created and more detailed models of the sites involved in the project are under construction. Currently the communications medium used for the project is the cellular mobile phone system thus allowing use when in range of a base station. In this particular project the planning and monitoring of civil engineering activities has been chosen as an exemplar for the technology, see Figure 1.

1.2 Background

Interdisciplinary in nature, this project draws together the work of three research groups from three departments within the University. These are the Transparent Telepresence Research Group (http://telepresence.dmem.strath.ac.uk) within the Department of Design Manufacture and Engineering Management, the Virtual Construction Simulation Research Group (http://www.strath.ac.uk/Departments/Civeng/conman/visrg.html) within the Department of Civil Engineering, and the

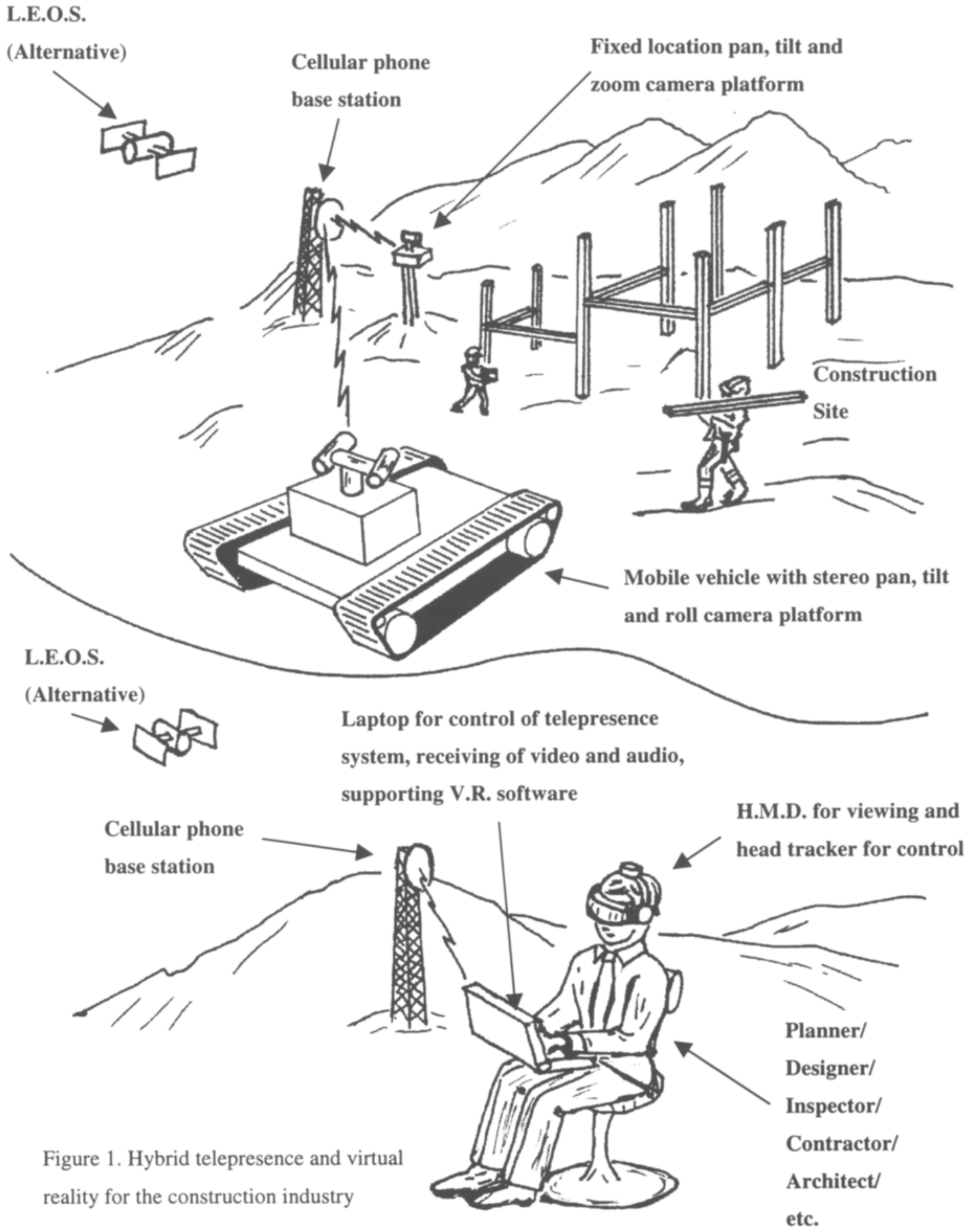

L.E.O.S.
(Alternative)
Cellular phone
base station
Fixed location pan, tilt and
zoom camera platform
Construction
Site
Mobile vehicle with stereo pan, tilt
and roll camera platform
L.E.O.S.
(Alternative)
Laptop for control of telepresence
system, receiving of video and audio,
supporting V.R. software
Cellular phone
base station
H.M.D. for viewing and
head tracker for control
Planner/
Designer/
Inspector/
Contractor/
Architect/
etc.

Figure 1. Hybrid telepresence and virtual reality for the construction industry

Compression Research Group (http://www.cs.strath.ac.uk/research/COMPRESS/index.html) within the Department of Computer Science. This matrix structure is described more fully elsewhere [1].

2. THE SYSTEM ELEMENTS

Within the following notes the "remote site" refers to the construction site or field location, the "home site" refers to the office, or mobile location such as a car, of the system user. For clarity of description the system can be considered as being composed of six major modules, see Figure 2, each of these being composed of their own respective subsystems. These modules are, the remote site hardware, home site hardware, control system, communications system, compression software, and the virtual reality module. They are all combined to create the hybrid VR telepresence system for simultaneous viewing of planned and actual real world progress. The modules utilise common equipment and hardware and software design concepts wherever possible, this ensures the full economic and functional benefits of standardisation are gained.

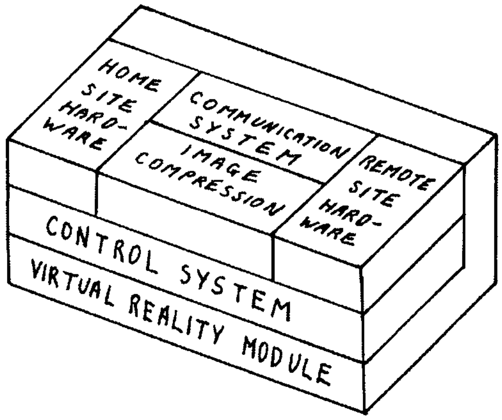


Figure 2. The integrated hybrid VR-TP system showing the interrelationship of each module

The system specification was thoroughly investigated before commencement of the detail design. A full product design specification (PDS) was created. One priority that was adopted was that the system should be small, portable and user friendly.

2.1. The Remote Site Hardware

The remote site hardware consists of a camera platform and support mechanism. Previously the telepresence research group have demonstrated their anthropomorphic sensor platform and described its design and use in telepresence systems [2,]. However the design of the camera platforms for this project will be non-anthropomorphic although they will still provide a number of degrees of freedom. One platform, for immersive telepresence and capable of pan, tilt, and roll, will have twin cameras for stereo viewing and be mounted on a tracked mobile vehicle. The vehicle is shown being tested with an earlier anthropomorphic head in Figure 3. The second platform, capable of pan and tilt, will have monoscopic vision with zoom capability and it will be mounted on a static platform. Communication hardware in the form of a mobile digital phone and interfacing card and control and drive hardware coupled with mobile power supplies are also required at this site.

The camera platforms utilise stepper motors controlled from a PC using a simple digital I/O (Input / Output) board. This I/O board can also be used to control other functions of the system, such as the zoom, iris and focus of the motorised lens of the camera.

The core of the hardware used in the telepresence system is a standard PC based around a laptop chipset. It incorporates one PCI expansion slot and a PC/104 expansion bus. The I/O board uses the PC/104 bus and the frame grabber used in the video system uses the PCI bus. The video information is taken from the frame grabber and compressed, this is then sent in one direction over the mobile phones. The camera platform and vehicle receive movement instructions again via the mobile phones from the home site controls, e.g. head tracker.

The physical construction of the remote site hardware for the camera platform is very small. The whole system has been designed to fit inside a box 250x250x200mm, this includes controls, communications, and battery power supplies. The unit is sealed and watertight and so can be used in outdoor conditions. It is also lightweight and robust allowing it to be moved around freely as the application demands.

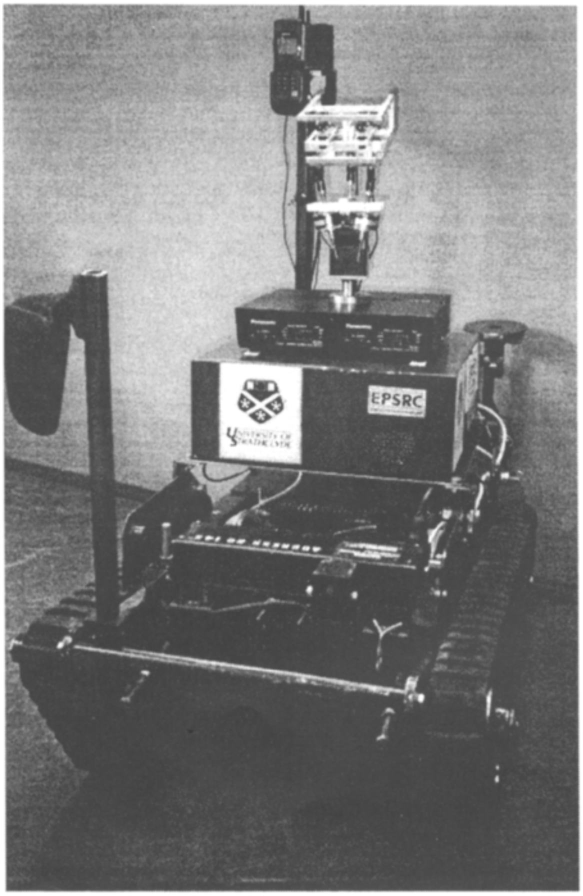


Figure 3. Mobile telepresence vehicle with anthropomorphic head

Much of the emphasis in design and construction has come from the desire to produce a system that could be commercially viable. The use of a standard PC at the heart of the system has provided the functionality to reduce costs in a variety of ways. The PC provides a reliable, proven platform that is inexpensive and ubiquitous. The complex hardware that is necessary in other telepresence systems has been moved to software providing a cheaper and more flexible system. Since standard PC components are readily available and inexpensive, upgrading the system is a very simple matter. The system can easily be tailored to the desired application.

Control of the stepper motors directly from the PC has also reduced costs by allowing simple hardware to be used. The problem of the timing complexity required to drive the motors has been transferred to software. The control solution adopted has proved reliable and accurate and still leaves most of the processor time free. This is necessary to grab and compress the video information in parallel. The control solution is also extensible to provide control for as many stepper motors as required at the remote site. The number is limited only to the processing power available. Thus as the power of PCs increase so the functionality of this system will also increase.

2.2. The Home Site Hardware

At the home site the minimum hardware requirements for non-immersive telepresence are a laptop PC with a modem and mobile phone. With this minimum configuration the remote vehicle and sensor platform can be controlled via the laptop's keyboard and the video from the remote site can be viewed on the laptop's screen. To provide immersive telepresence the decompressed video is sent to a head mounted display with an integrated head tracking device. The entire home site configuration using the laptop PC is, like the remote site, battery powered thus allowing total mobility.

2.3. The Control Software System

The software control system allows control of camera zoom, iris and focus, the camera platform pan, roll and tilt, and vehicle forward or reverse, steering and speed control. In common with the hardware described above the software is based on standard off-the-shelf PC operating systems and tools. Having evaluated several operating systems, we selected FreeBSD, an advanced BSD UNIX operating system for PC compatible computers. There were many reasons for this decision. Full source code for FreeBSD is available, this eases the task of systems and device driver programming by allowing new programs to be created based on the experience gained from other proven programs. There is also good support for this software including direct contact with the authors of the system. Communication facilities within FreeBSD are excellent thus facilitating the remote administration that is essential for the intended applications of our system. There is reasonable support for the many inexpensive and commonplace peripherals available for standard PC components, this is essential to realise the full benefit of these devices. Having available the source code to the drivers allows changes specific to the needs of the project to be easily made.

FreeBSD also has excellent multitasking capabilities, essential since the system is expected to do many things simultaneously, such as compressing video, moving motors, and managing communications. Resource usage is also efficient, allowing us to specify a minimal system leading to lower power consumption and reduced mass and space requirements. Fast development of new device drivers, such as those required by the motor driving subsystem, is aided by the availability of "loadable kernel modules". These allow changes to the system to be evaluated without the need to reboot the system each time.

2.4. The Communication System

A unique aspect of the system is the use of mobile phones for communications to provide maximum mobility and ease of access. This is one of the reasons we have chosen to base the communication system around the internet protocol TCP/IP which has some excellent methods of communicating via mobile and standard phones. One example of this is PPP (point to point protocol) and a multilink version of this, MPPP, which allows multiplexing of individual modems to increase bandwidth.

Another advantage of using TCP/IP is that it is undoubtedly the most common networking protocol in use today. This means that, should we choose to do so, we can easily make our systems accessible to the majority of the world without modification. It also allows us to use the communication software already present in the system at an abstract level - the system is in general unaware of whether or not it is connected to, for example, a fast local Ethernet, ISDN, ATM, or mobile phones since the underlying mechanisms used to communicate are exactly the same. Thus our system is designed to be used with many different communications settings and bandwidths, without modification.

2.5. Image Compression

When there is just one camera at the remote site, there is just one video stream to compress. A standard video compression CODEC can be used providing it meets with our target bandwidth of 9.6 kbps and runs in real time. Ideally it should be a software CODEC minimising the required hardware. Our options were the ITU standardised H.263 CODEC or the University of Strathclyde's SCT (Strathclyde Compression Transform) CODEC since the MPEG CODECs are only suited to much higher bandwidths.

The SCT [3] offers many advantages over H.263. It is smoother on low bandwidth links as it offers approximately a constant number of bits per frame. It is easily extended with new encoding methods and its low algorithmic complexity allows it to run in real time on low power processors (e.g. Pentium 120) or deliver better quality than H.263. As the CODEC was developed by the Department of Computer Science we had an intimate knowledge of the compressor's implementation.

When two cameras are used to form a stereoscopic system, there are two video streams to be compressed. A simple approach would be to use two compressors, one each for the left and right video streams, both working independently. This is not the optimum approach. A specially written stereoscopic compressor can produce a better quality image by utilising the large degree of similarity between images from the right and left cameras.

If both cameras are mounted on a horizontal platform, at a simplistic level, objects in the left image are the same as those in the right image but with a horizontal offset. Due to occlusion one camera may see an object the other camera cannot and this must be considered in the design of the compressor. Efficient stereoscopic compression can then be performed by compressing the left image using SCT (or any other monoscopic compressor) and compressing the right image by storing only these horizontal offsets. A thorough literature search has revealed the uniqueness of the real time low bitrate stereoscopic compression being developed here.

2.6. Virtual Reality Module

The VR module allows the creation of a "virtual construction project" based on a project schedule and using Superscape. It also allows subsequent visual monitoring of, and interaction with, the progress of the simulated project with respect to spatial and temporal variables. It is based on the prototype developed at Strathclyde's Virtual Construction Simulation Research Group (VCSRG) [4]. The module is composed of two main elements; the first is a generic library of the 3D images of the project site, building structures, environment, and activities; and the second is a means of subsequent visual simulation of the construction process. Operations carried out by the module include the following. (a) Plan formulation through generation of construction and site related activities and their sequence, schedule generation and resource allocation etc. (b) Manual and automated mapping of activities and facilities to their appropriate graphical representations. (c) The projection and simulation of the progress of work in a virtual environment based on the project schedules.

3. HYBRID VR AND TELEPRESENCE

As noted earlier the construction industry is taken as an example application for this technology. The complexity of the operations necessary to construct complex buildings or civil engineering structures ensures that efficient monitoring and control are key management issues. The gathering of site-specific information and its integration with current and planned activities is therefore essential. All of this indicates the need for a means of integrating the information contained in the virtual model outlined above, with the real world images and data gathered by the telepresence system.

This synthesis we have termed "hybrid VR and telepresence" to differentiate it from the commonly used term "augmented reality". We have done this for two reasons. Firstly in our system we are incorporating images gathered at a distance therefore we wish to emphasise the telepresence aspect. Secondly, in augmented reality the real world image is overlaid with virtual elements, or the virtual world is overlaid with real elements. In our case with the time and resources available for the project we are under no illusion as to the difficulty of obtaining the necessary registration in space and time, occlusion difficulties, scale considerations, and other difficulties to be resolved in real time. We therefore expect at this stage to achieve a rudimentary overlay of images, possibly with a side by side visual comparison being more effective. for the user.

4. CONCLUSION

At the time of writing the project is two thirds of the way through its two year duration. Successful preliminary demonstrations of the telepresence, image compression, and mobile phone systems have already been carried out [5]. The telepresence equipment is now being deployed on an actual building site that is also being modelled using the VR module.

The possibility of being able to access all of the facilities remotely is extremely attractive to potential users. The accessibility will be enhanced with the introduction of Low Earth Orbit Satellites (LEOS) which will allow communication from anywhere in the world using mobile telephone handsets. This should enhance the systems potential further since the possibility of increased wireless bandwidth is also probable.

ACKNOWLEDGEMENTS

The EPSRC MNA Programme (Grant GR/L06164) is supporting this work in collaboration with Alvis Logistics, Babtie Engineering, and Orange Personal Communications.

REFERENCES

[1] G. Mair, Telepresence - The Technology And Its Economic And Social Implications, Proc. IEEE and IEE International Symposium on Technology and Society 1997, ISBN 0-7803-3982-7

[2] G. Mair, J. Heng, R. Fryer, N. Clark, D. Sheat, An Experiment In Long Distance Telepresence Using ISDN2, Proc. Mechatronics '96, Guimaraes, Portugal, ISBN 972-8063-08-3

[3] R. Lambert, R. Fryer, W. Cockshott, D. McGregor, Low Bandwidth Video Compression with Variable Dimension Vector Quantisation, Proc. ADViCE '96, Cambridge, UK, July 1966

[4] A. Retik, VR System Prototype for Visual Simulation of the Construction Process, EPSRC Conference, Salford, 1995.

[5] "Mobile Virtual Reality Is Combined With Telepresence" IMPACT (EPSRC Newsletter), March 1998.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Development of an electropneumatic device for remotely sensed manipulation

C. Ferraresi, A. Manuello Bertetto, H. R. Ziarati Niasar

Department of Mechanics, Technical University Politecnico di Torino
C.so Duca degli Abruzzi, 24 - 10129 Torino - Italy

This paper provides a method to develop by computer simulation a device for remotely sensed manipulation, assembled with low-cost pneumatic components and sensors.
The methodology for an effective identification and modelization of the components is also described.
The non-linear overall model has been implemented and simulated in Matlab-Simulink environment, in order to individuate the gain values of the control algorithm that, although simply proportional, led to satisfactory results.

1. INTRODUCTION

The problem of remote manipulation of objects may arise in various situations where a human being cannot perform directly the operation. In this case the operator will supply manually to a proper sensorized device the reference of the gripping force. Such device, that could be defined as a "sensorized glove", can in addition be equipped with actuators, controlled in position, that give to the operator the sensation of stiffness presented by the object during the manipulation.

The subject of remotely sensed manipulation has been studied by many researchers and a number of solutions have been proposed, mainly adopting electro-mechanical or hydraulic solutions for the actuations.

With respect to these classical solutions, a device based on the use of pneumatic actuators may present some advantages in terms of cost. In addition, the control of the gripper force may be obtained by the control of the air pressure in the correspondent actuator, quite easy to achieve [1].

On the other hand, some difficulty may arise for the position control of the "glove" if pneumatically actuated. However the motion control of pneumatic actuators has been studied since long time and even the authors of this paper proposed several solutions [2]. For all these reasons, a system for remote sensorized manipulation made up of low-cost pneumatic components and sensors has been considered.

The gripper and the glove control only one degree of freedom, assuming that the results obtained can be extended to more complex systems.

The aim of this work is to demonstrate the feasibility and effectiveness of such a system. The study is carried out by modelization and simulation in Matlab-Simulink environment.

Great importance is given to the identification and modelization of the various components, in particular the valves and the tactile sensors.

The modelization of a pneumatic system requires the identification of all their components; the accuracy of this phase has a great influence on the reliability of the simulations.

In the literature many researchers propose models of pneumatic systems both with lumped and distributed parameters.

In [3] the author presents a general method to modelize pneumatic circuits considering lumped parameters, describing the equations related to the several components: tubes, fittings, actuators and valves.

For fluid-mechanic systems subjected to quick transients, with rather long tubings, a lumped parameters modelization may be inadequate. In [4] a method for dynamic modeling and simulation of such systems is proposed, using the characteristics method for the integration of the partial differential equations of the model.

In the present work a lumped parameters model has been considered. This choice is fully justified since in the system studied the propagative effects are negligible, because of the limited length of the tubings. This led to a rather simple but accurate model, that allowed to take into account all the main physical characteristics and to individuate a feasible control method.

2. DESCRIPTION OF THE SYSTEM

An overall scheme of the system is shown in figure 1. The single action pneumatic cylinder C_1 is defined "gripper" and is devoted to grasp the object 1, controlling the force F_G. This latter is detected by the force sensor S_1 and compared with the reference force F_R, exerted by the operator and measured by the force sensor S_2.

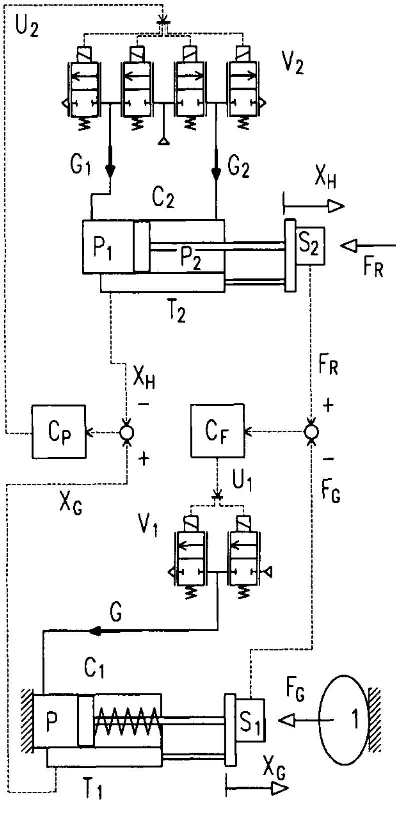


Figure 1. Scheme of the electropneumatic system.

The error F_R-F_G is compensated by the controller C_F which acts on flow-proportional valves V_1 and controls the force exerted by the cylinder C_1.

The force sensor S_2 is placed on the rod of the double action pneumatic cylinder C_2, whose position is controlled by the controller C_P through the flow-proportional valves V_2.

The input to the controller C_P is the error between the position x_G of the gripper, detected by the transducer T_1, and the position x_H of the cylinder C_2, detected by a second transducer T_2. In this way, the cylinder C_2, which is handled by the operator, complies like the grasped object 1.

Therefore the operator not only can grasp the object located at any distance and with a controlled force, but has also the feeling of the stiffness of the remote object. Each force sensor consists of a pneumatic microactuator connected to a pressure transducer.

This kind of sensor [5], although very cheap, proved to be quite accurate and prompt for this kind of application.

3. IDENTIFICATION AND MODELING

In order to modelize the whole system it is necessary to characterize the behavior of all its components.

A block scheme of the model is shown in figure 2.

The input to the system is the external force F_R, exerted by the operator and detected by the sensor S_2. The correspondent signal is compared with the one generated by the sensor S_1, which measures the force F_G exerted by the gripper. The force error E_F is processed by the controller C_F, which sends the command U_1 to the valves V_1.

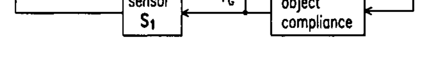


Figure 2. Block diagram of the system.

The flow rate G, together with the gripper force F_G, determines the position x_G of the gripper itself.

The block "object compliance" simulates this characteristic of the manipulated object.

The gripper position x_G is sensed by the transducer T_1, whose signal is compared with the one generated by the analogous sensor T_2, installed on the actuator C_2 (defined as the "glove"). The difference between the positions of gripper and glove is processed by the controller C_P, which sends the command U_2 to the valves V_2.

The position x_H of the glove is determined both by the air flow rates from the valves V_2 and by the operator force F_R.

For each component of the system a mathematical model has been derived. In all the following equations the symbols are the same indicated in figure 1 and described in nomenclature.

3.1. Force sensors

The force sensors included in the system are two tactile low cost pneumatic sensors [5], each one made up of a pneumatic membrane microactuator connected to a pressure transducer, as shown in figure 3.

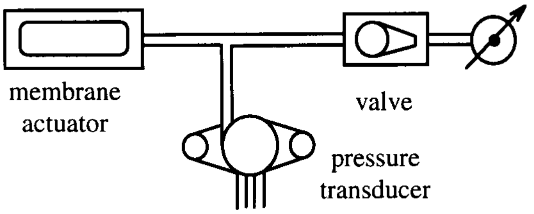


Figure 3. Scheme of the force tactile sensor.

A cross section of the membrane actuator is presented in figure 4. The actuator is a commercial membrane pneumatic cylinder FESTO EV-15/40-4, with a stroke of 4 mm and an equivalent bore of 28 mm; it is normally pre-loaded with a certain air pressure (0.1 bar).

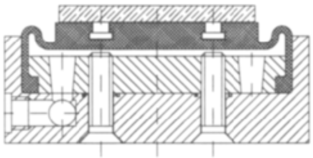


Figure 4. Cross section of the tactile sensor actuator.

The pressure transducer is a MOTOROLA 2010GP. When an external force acts on the actuator membrane, the internal pressure raises, following an almost linear static characteristic. Experimental tests proved that the dynamic behavior of the sensor can be effectively assimilated to that of a linear second order system. Therefore the relation between the microcylinder pressure and the external force is:

$$\frac{p}{F_R} = \frac{K\omega_n^{\,2}}{s^2 + 2\zeta\omega_n s + \omega_n^{\,2}} \tag{1}$$

where K is the static gain, ζ the damping factor and ω_n the natural angular frequency. By recording the sensor response to reference force steps, it has been possible to evaluate the various parameters: ζ=0.03; ω_n =2500 rad/s; K=6·10^4 Pa/N.

3.2. Pneumatic actuators

The gripper is actuated by a single action pneumatic cylinder, which is normally retracted because, once the manipulation is concluded and the glove is not operated, the gripper and the glove itself must remain open in a defined position. On the contrary, the glove is actuated by a double action cylinder.

For the continuity of the air flow-rate entering the cylinder C_1 one can write:

$$G = \frac{d(\rho W)}{dt} = W\frac{d\rho}{dt} + \rho\frac{dW}{dt} \tag{2}$$

Assuming the air as a perfect gas and considering isothermal transformations, it is possible to express the pressure derivative:

$$\frac{dP}{dt} = \frac{\dfrac{GP_0}{\rho_0 A} - P\dfrac{dx_G}{dt}}{x_G + x_m} \tag{3}$$

and analogously for the pressures in cylinder C_2:

$$\frac{dP_1}{dt} = \frac{\dfrac{G_1 P_0}{\rho_0 A_1} - P_1\dfrac{dx_H}{dt}}{x_H + x_m} \tag{4}$$

$$\frac{dP_2}{dt} = \frac{\dfrac{G_2 P_0}{\rho_0 A_2} - P_2\dfrac{dx_H}{dt}}{-x_H + x_m + x_0} \tag{5}$$

In equations (3), (4) and (5) the air flow-rate is considered positive when entering, the subscript 0 in pressure and density refers to ambient conditions.

For the cylinder C_1 the force equilibrium can be expressed as:

$$m_G\ddot{x}_G + F_{F1} + k_G(x_G + x_{G0}) + F_G = \\ = A_{1G}(P - P_0) - A_{2G}P_0 \qquad (6)$$

and analogously for the cylinder C_2:

$$m_H\ddot{x}_H + F_{F2} + F_R + (P_2 - P_0)A_2 = A_1(P_1 - P_0) \qquad (7)$$

In equation (6) the data referred to the return spring (k_G and x_{G0}) have been measured directly. The friction forces (F_{F1} and F_{F2}) have been deduced from the literature [6] for cylinders of this type and dimensions.

3.3. Pneumatic valves

All the six valves used are two-way digital valves MATRIX 821, controlled with PWM signals to obtain a modulation of the air flow rate. Therefore it has been necessary to determine their flow characteristic. This has been made according to the ISO 6358 [7], by measuring the conductance and the critical pressure ratio of one valve activated by a constant signal. In this condition the valve has a conductance of $2.4 \cdot 10^{-9}$ Nm³/sPa and a critical ratio equal to 0.48.

The model assumes that the conductance of the valve during the PWM modulation is proportional to the duty-cycle of the corresponding PWM signal generated by the controller, while the critical ratio maintains the value measured at full-open condition. In order to verify that, several measurements of the actual flow-rate at various duty-cycle values and at different modulation frequency have been carried out. Figure 5 describes the results of these tests: the curves show the value of the valve conductance versus the signal duty-cycle, for three different modulation frequencies (10, 20 and 50 Hz). As one can see, the conductance has an almost linear trend for the frequencies considered.

Therefore, the assumption of considering the conductance proportional to the signal duty-cycle is verified if the signal frequency is 50 Hz or less. The tests showed also that the critical pressure ratio of the valve may be considered almost constant at the various operating conditions.

As shown in figure 1, for each cylinder chamber two valves are needed: one for supplying and one for exhausting. Looking for example at valves V_1, controlling the flow-rate G to the actuator C_1, when a positive G is needed, the control C_F must activate the right valve, connecting the actuator to the supply, and close the left one, and vice-versa for a negative G.

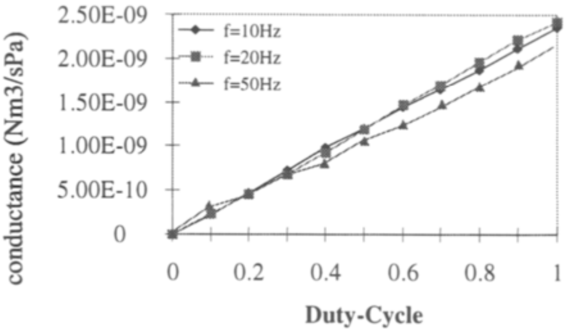


Figure 5. Valve conductance versus the duty-cycle of the PWM signal.

The mass flow-rate G can be expressed as per the ISO 6358:

$$G = K_t\rho_n C P_U \sqrt{1 - \left(\frac{(P_D / P_U)}{1 - b}\right)} \quad \text{for } P_D/P_U > b$$

$$G = K_t\rho_n C P_U \qquad \text{for } P_D/P_U < b \qquad (8)$$

ρ_n is the air density at standard reference conditions (293.15 K and 10^5 Pa)

In equation (8) the upstream pressure P_U may correspond either to the supply pressure or to the pressure P in the actuator, while the downstream pressure P_D may correspond either to the actuator pressure P or to the atmospheric pressure P_O, depending on the operating condition of the valves. The conductance C is expressed by the full open value multiplied by the value of the duty-cycle of the PWM command signal. The equation (8) refers to the flow-rate entering the cylinder C_1. The flow-rates G_1 and G_2 related to the actuator C_2 may be expressed in the same way.

4. SIMULATION

Once defined all the physical parameters of the components, the behavior of the whole system has been simulated, in order to individuate the parameters of the force and position controls and verify the effectiveness of the system.

The complete system model is again the one shown in figure 2; the block "object compliance" is necessary to simulate the reaction force of the grasped object. For both the gripper force and the glove position simple proportional control algorithms have been adopted.

The input to the model is the reference force F_R exerted by the operator in the real system. A number of simulations have been made, with different values of the force gain K_F and the position gain K_P. The results obtained for the optimal gain values, which are: $K_P=100\ m^{-1}$ and $K_F=0.05\ N^{-1}$, are shown in figures 6, 7, 8 and 9. These simulations were performed considering an elastic object, with a stiffness of 4000 N/m.

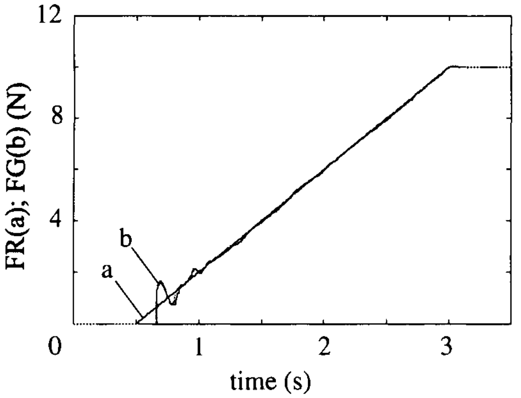


Figure 6. Reference force (a) and simulated gripper force (b).

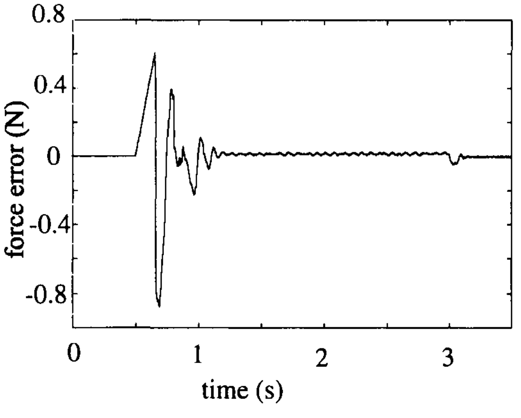


Figure 7. Force error.

Figure 6 shows the outline of the reference force and the gripper force. Figure 7 shows the correspondent force error: its maximum value is around 0.8 N and becomes almost zero in about half a second. The relatively high error in this first phase is due to the acceleration of the gripper before it comes in contact with the object: this determines an impact force that depends on the length and velocity of the gripper free stroke.

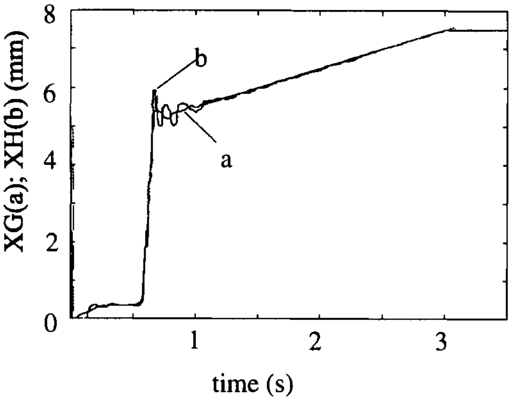


Figure 8. Simulation of gripper position (a) and glove position (b).

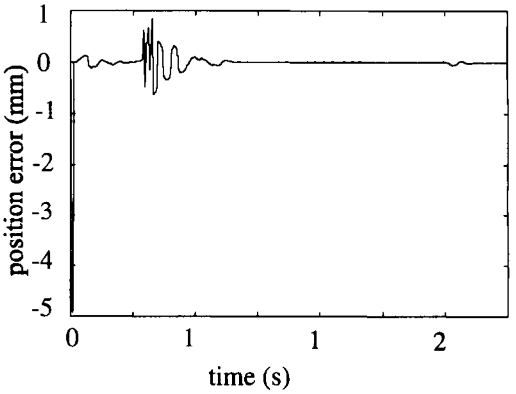


Figure 9. Position error.

Figure 8 shows the outline of the gripper position and the glove position. Figure 9 shows the correspondent position error: its maximum value is about 0.8 mm, but reduces drastically in quite a short time.

From zero to 0.5 s the force input to the model is zero, but the model is active and moves to get an equilibrium condition, under the effect of the supply pressure.

When the force input begins to vary, both the actuators perform very fast and synchronous motions, until the gripper comes in contact with the object. At this time, the glove position and the gripper force show their maximum dynamic errors.

However, these latter may be reduced by adopting proper pneumatic resistances for the air flow-rates to the actuators, in order to increase the overall damping, or considering control laws more effective than the simple proportional here adopted.

CONCLUSIONS

This paper presented the method to design by computer simulation a device for remotely sensed manipulation based on the use of pneumatic components. The methodology for an effective identification and modelization of the components has also been presented. The particular characteristics of all the components led to a highly non-linear model, whose behaviour has been simulated in Matlab-Simulink environment. The simulations allowed to individuate the gain values of a simple proportional control able to achieve quite satisfactory results for a remotely sensed manipulation. Improving of the performance could be obtained both by considering more effective control laws and adopting slight modifications of the pneumatic circuit. These solutions may easily be introduced in the model and their effect directly evaluated.

This study demonstrated the feasibility and effectiveness of the system, although assembled with very low-cost components and sensors.

REFERENCES

[1] Ferraresi C., Sorli M., Manuello Bertetto A., Pastorelli S., "Force Control and Grippers", Congresso Internazionale della Trasmissione di Potenza, Milano, 20-21 June 1995.

[2] Ferraresi C., Giraudo P., Quaglia G., “Non-conventional adaptive control of a servopneumatic unit for vertical load positioning”, Fluid Power West Technical Conference, Anaheim, USA, 23-24 March 1994, pp. 319-333.

[3] Ferraresi C., “Modellazione di sistemi pneumatici”, Fluid n.346, pp. 48-54, February 1993.

[4] Romiti A., Raparelli T.: “Dynamic Modelling and Simulation of Pneumatic Systems”, Fluid Power - Proc. IX Int. Symp., 1990, Cambridge, U.K. pp. 137-146.

[5] Ferraresi C., Manuello Bertetto A., Mattiazzo G.: “A Low Cost Pneumo-Electronic Tactile Sensor”, Int. Conf. on Recent Advances in Mechatronics, Istanbul, Turkey, pp. 1098-1103, 1995.

[6] Belforte G., D’Alfio N., Raparelli T.: “Experimental Analysis of Friction Forces in Pneumatic Cylinders”, The Journal of Fluid Control, Vol. 20, n.1, Stanford, Nov., 1989.

[7] ISO 6358 - Pneumatic fluid power - Components using compressible fluids - Determination of flow-rate characteristics.

NOMENCLATURE

A	cylinder chamber thrust area
b	critical pressure ratio
C	pneumatic conductance
F_F	cylinder friction force
F_G	force on the object (from the gripper)
F_R	reference force (from the operator)
G	air mass flow-rate
K	tactile sensor static gain
k_G	spring stiffness
K_T	temperature coefficient
m	mobile mass of the actuator
p	relative air pressure
P	absolute air pressure
W	cylinder chamber volume
x_0	stroke of cylinder C_2 (glove)
x_{G0}	spring pre-load deformation
x_m	cylinder chamber dead space
ζ	tactile sensor damping factor
ρ	air density
ω_n	tactile sensor natural angular frequency

ACKNOWLEDGEMENTS

This work has been supported by the Italian Ministry of University and Research (MURST).

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

A Mechatronic Approach to Auditory Localisation for Transparent Telepresence

Colin S. Harrison B.Sc. (Eng. Sci), M.Sc.
email: colin@vulture.dmem.strath.ac.uk

Gordon M. Mair B.Sc., DMS, MIEE, C.Eng.
email: g.m.mair@strath.ac.uk

Transparent Telepresence Group, Department of Design Manufacture and Engineering Management, James Weir Building, 75 Montrose St, Glasgow, G1 1XJ.

ABSTRACT

This paper describes the importance of the sense of hearing for immersive telepresence systems, and a fundamental approach to designing and implementing a binaural sensor system. This approach relies on a knowledge of how the human auditory localisation mechanisms work to locate sounds in space in order that a synthetic system can be derived. The most important of these mechanisms are explained. The Strathclyde system is then described including ideas drawn from the literature to define a binaural approach. The system consists of a set of artificial pinnae and microphones mounted on a binocular stereo sensor platform capable of pan, tilt and roll which is slaved to a head tracking system. A set of preliminary experiments are described which test the performance of the system in practice as a basis for further development.

1. INTRODUCTION

We define Transparent Telepresence as the use of technologically mediated experience to enable a telepresence system operator at a home site to experience being fully present at a remote site. This implies that the intervening technology will be totally transparent to the user. A full system would therefore allow the user to experience the remote site via all five classical senses plus vestibular and kinaesthetic sensations. The creation of such a transparent system lies decades in the future. However at the "Transparent Telepresence Group" within the Department of DMEM a telepresence system has already been developed [1] incorporating a mechatronic stereo camera platform, communication links and immersive stereopsis display equipment. To augment this system we are developing a binaural sound capability to greatly enhance the users' sense of presence. The importance of hearing in contributing to a sense of presence is well documented. For example in a study of suddenly deafened adults it was noted that they experienced a radically reduced sense of reality and presence in the world around them.[2] Relatively little attention has thus far been given in the literature to the potential of sound for telepresence compared with vision systems. Analogous work is however going on in virtual systems to enhance the sense of presence in computer generated environments and in cockpit information systems for military aircraft to provide directional information to the pilot on approaching aircraft or missiles.[3] By incorporating binaural sound into the system we are able to enhance the level of realism generated by the remote environment and presented to the user. This relies on a knowledge of the processes which humans use to localise and externalise sound[4]. This paper describes some of these mechanisms and some preliminary tests used on the mechatronic telepresence head.

2. HEARING MECHANISMS

2.1 Interaural Timing Differences

Humans are able to determine where a particular noise is coming from based on a series of complementary mechanisms. Understanding of how these mechanisms work gives the basis for a design which can then be used to generate a telepresence aural detector. The objective is to define a system such that the user cannot tell whether they are listening through their own ears, or through the Transparent Telepresence system. The main mechanisms involved in auditory localisation are based on the fact that we have two ears placed on opposite sides of a moveable head. The mechanisms are based on differences between the ears in time (or phase), in intensity, in multiple sampling by moving the head, and on the presence of a convoluted cartilage cup, the pinnae which generate complex distortions in the upper frequency ranges. The whole assembly is mounted on a roughly spherical container which generates its own effects, the head. At different frequencies different mechanisms become more important, in other words, how you hear depends on what you hear. Assuming a nearly plane wave front the distance travelled by the further wave element can be shown to be dependent on the diameter of an idealised sphere and the angle made with the sound source.[5]

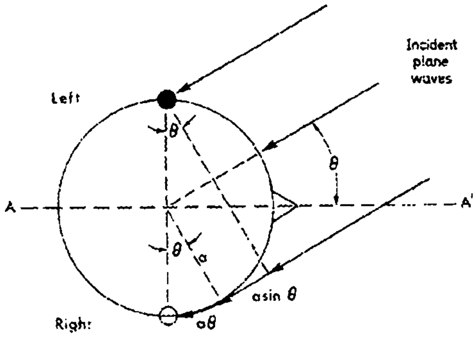


Figure 1 Interaural Time Differences

For radius a= 8.75 cm and assuming a speed of sound c= 343 m/s then the interaural time difference is given by:

$$\Delta t = \Delta a / c = 255(\Theta + \mathrm{Sin}\Theta)$$

where the time t is in microseconds.

For frequencies below about 1500 Hz this formula agrees well with experimental results. Provided the time differences do not overlap then these differences can specify a position in Azimuth for the source of a sound.

2.2 Interaural Intensity Differences

The differences in intensity are not so well behaved, and rely on the fact that the far ear is in a sound shadow whose shading by the head depends on the frequency of sound and its location in space The interaural differences define the position of a sound on a possible cone but are not enough to uniquely define the position in space.[3]

2.3 Moving Head

The interaural differences in time and intensity based on stationary holes on each side of a stationary head are not sufficient to uniquely specify the position of a real source of sound. However the ability to move the head and therefore to repeatedly sample from the same source improves the ability to pinpoint the sound source. This helps to remove some of the ambiguity in human auditory localisation, particularly front-to-back errors.[5] A front-to-back error is where a sound which is correctly located in the front is perceived by the listener to be in the rear. Back to front errors also occur but are much less common.

2.4 Pinnae

The pinnae and head form an acoustic filter which is unique to each of us. Individual ears differ and The distortions that they introduce into the upper frequencies at the impact with the head/ear combination differ for each of us. This effectively prevents the development of a truly unique physical model of the head which will give a transparent solution for all of us. This is a limitation of the binaural approach. Nevertheless it is possible to specify general or median approaches and good results are obtainable. [6][7]

2.5 Distance Perception

The distance at which we perceive an object is governed by several factors, though the detail of the mechanisms seems to be complicated and not well defined.[3][8] The perceived distance of a sound is only weakly dependent on intensity. Other factors include the ratio of direct to reverberant sound. Dispersion effects on the frequency content are also important as the higher frequencies tend to be more absorbed by moisture in the air. Additionally the pattern of the early reflections from surfaces in the environment provide important cues to distance.

2.6 Externalisation

Externalisation in the audio context refers to the position of an audio image relative to the listeners own head. Lack of externalisation is the phenomena where a perceived audio image is located inside the head. This is a common experience with the conventional intensity stereo recordings that we have at home when we listen over headphones. Of course stereo was not designed to be listened to over headphones, it is just a handy way of listening without disturbing the neighbours. A feature of listening through loudspeakers is that the sounds are automatically externalised as the sound truly emanates from outside the head. Hartmann[9] has shown that with very careful attention to the interaural phase and level differences realistic signals indistinguishable from real world sources can be generated. From the view of the telepresence application then it will be seen that it is possible, based on practical evidence to generate an artificial signal technically indistinguishable form a real world signal (the baseline criterion) that fulfills the criteria that sound should be compact, externalised and in the correct location in space.

2.7 Visual Cues

The objective of this system is to provide the most human like audio system. In practice however it should be noted that eyes can trick ears, and if contradictory cues are supplied to the vision and aural senses then often the visual sense overrides the human perception. Wallach[10] describes an experiment where rotating stripes fool the eyes into believing that the observer is in motion and this produces the audio illusion that sounds correctly located ahead are reported as above.

3.0 AUDIO SYSTEMS FOR TELEPRESENCE

The system that was chosen for this particular preliminary installation was based on Binaural Sound. Common implementations of this system use dummy heads with artificial pinnae attached to microphones at the ear positions. It is possible to generate quite realistic sounds using binaural recording techniques with reproduction working best with headphones. However individual physical differences mean that a truly universal head based on physical dimensions alone is not possible. Instead a median male head such as the KEMAR [7] head is used which uses the averages from a set of males. Several recordings have been made using this technique and CD's are available which the interested listener is invited to audition. Binaural recording has made several attempts to break through into the mass market but has never made the jump to mass popularity in the recording field as the results work best for headphones and most music listening is apparently over loudspeakers.

Thus a binaural system was chosen for the Strathclyde head, and was designed bearing in mind the "a priori" structure containing a Stereo vision cameras and the ability to move the head under slave control in pan, tilt and roll.

This uses two microphones placed at the analogue of the human head position on the robotic head, thus fulfilling the anthropomorphic and anthropometric requirements. A form of design incorporating these requirements was therefore envisioned which would consist of two microphones, two small microphone pre-amplifiers, and artificial pinnae. The idea that an ear/microphone based system could be used was also derived independently from work on Auditory Localization done by Batteau.[11] Figure 2 shows the results of an experiment performed by Batteau which tested the localization accuracy of a pinnae microphone system in the azimuth plane (i.e. parallel to the surface of the earth. A pair of maracas were shaken at different positions around the apparatus and the listener then asked to indicate the perceived position of the sound which was detected remotely over headphones. A good correlation can be seen between actual and reported azimuth.

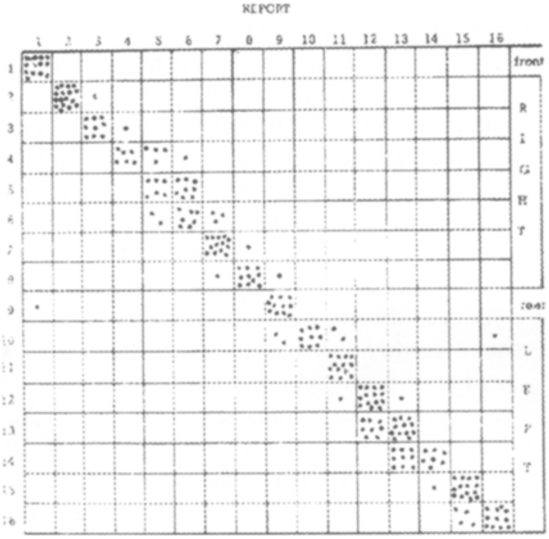


Figure 2 Correlation for reported and actual position of sound detected remotely.[11]

3.1 Experimental System

The experimental system for binaural telepresence consisted of the following main parts.

External Pinnae (cast from the authors own ears)
Adjustable Mounting for the pinnae
Microphone Amplifiers
Polhemus head tracking system + Cybereye Virtual Eye glasses
Associated Computers for Head motion + Control Software
Strathclyde Mechatronic Telepresence Head.

The External Pinnae were cast from the author's own ears by The National Prosthetic Centre. The casting included the concha but not the ear canal (auditory meatus) as this was synthesised by using a cylindrical chamber within the mounting system. Some experimental arrangements use an ear canal and some do not, therefore it was decided to make this a feature which could be varied by the physical arrangement. Additionally the ellipsoidal structure of the human ear means that the alginate was expected to break off inside the head at the casting stage.

The mounting system was based on the requirement to have a physical ear canal, adjustability in the x and z direction and the ears could also be rotated in the z plane. The system was also to be designed within a cash spend excluding manhours of about £100.

3.2 Experiments

A number of experiments were performed in the University of Strathclyde anechoic chamber in order to test if the system worked at all and if so which factors were important. The tests were based on the user's ability to discriminate in the azimuth plane only for the following arrangements.

Using the pinnae without head motion
Using the pinnae with head motion
No pinnae with head motion
No pinnae without head motion

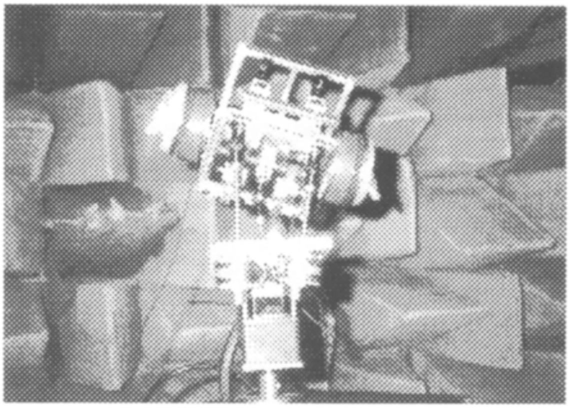


Figure 3 Audio for Telepresence Test Apparatus

Vision cues were not used in the tests in order to ensure that the audio cues were considered in isolation.

The user sat outside the University anechoic chamber in order that the physical location of the mechatronic head could not be seen. The remote head system was placed on a rotary table in the anechoic chamber and then the user attempted to locate the source of the sound using the telepresent head and the ear assembly. This was compared against a 12 position grid corresponding to the positions on a clock face. This was then repeated 18 times. The subject was allowed to move the head within the motion parameters of the Strathclyde head. This allows for movement within 158 degrees of Azimuth and 121 degrees of pitch and 82 degrees of roll. This mimics the head motion in a natural way.

3.3 Experimental Results and Discussion

	R	E	P		A	Z	I	M	U	T	H		
A		1	2	3	4	5	6	7	8	9	10	11	12
C	1	•	•										
T	2	•	•										
	3				••								
A	4					•							
Z	5					•							
I	6						•	•					
M	7							••					
U	8								•				
T	9									•			
H	10										•		
	11											•	•
	12												•

Figure 4 Head motion + pinnae

It will be seen from the summary test results that the general pattern expected in the literature survey is mirrored by the results. The differences in outcome are mainly those of emphasis. The major caveat to all the results is that a very low number of trials was conducted and for only two subjects. Nevertheless the repeating patterns are evident particularly as a first estimate of how further tests might be conducted.

It will be seen from these basic tests that the main factors governing the ability to localise are:
Head motion (vs. no head motion)
Having artificial pinnae

	R	E	P		A	Z	I	M	U	T	H		
A		1	2	3	4	5	6	7	8	9	10	11	12
C	1				•	•							
T	2		•	•									
	3		•										
A	4				•								
Z	5	•		•									
I	6					•							
M	7							•					•
U	8								•				
T	9									•			
H	10								•	•			
	11						•						
	12					•							•

Figure 5 No Head motion and no pinnae

Whilst many more trials are necessary to be certain, the addition of ears, particularly our own, will help to reduce the number of front to back confusions, even without the addition of vision cues, and the accuracy of the ability to localise is also increased. Listening through the ears of other people has apparently, for one subject at least, less impact on the ability to localise. This may pose problems as in telepresence a general solution is desirable and it may not be possible to individually cast ears and substitute them for each individual user. The ability to localise through other peoples ears (in fact other peoples complete head-ear-torso combinations) has been investigated by Wenzel [3] for the NASA AMES project whose conclusions were that provided HRTF's from "good localisers" were used then other good localisers obtained reasonable results, though not as good as their individualised HRTF's. Involving bad localisers in the procedure either as a model for the HRTF's or listening through good HRTF's produced little improvement. This suggests that there are characteristics that could be generalised to form an average ear (HRTF) that will generate better results than having no ears at all. The ideal solution would be a programmable ear which could be developed to mimic individual characteristics of the user at the time. This would mean setting an individualised hearing profile for each telepresent user in order to optimise the system and to produce a realistic sound for that specific user. In fact this is a standard feature of the system already in that, in relation to vision, the pupillary distances and focus are set for each individual, and the head mounted display is adjusted for head diameter. It represents another variable that must be personally set for each user.

4.0 CONCLUSIONS

An initial system for auditory localisation in telepresence has been successfully developed within the project constraints. A series of limited trials has been conducted from which tentative conclusions may be drawn. The implementation of a basic microphone based system with two separate channels for left and right ears, and head motion provides a good means of localising sound remotely, using a telepresent head. The model used was that known as binaural recording, from which further developments may be made. The addition of ears, particularly those cast from the users ears, improves the ability to localise and reduces the number of gross errors, in one case to zero. Where the ears used are not those of the user, then they seem to confer a marginal advantage.

The work by Batteau [10] suggests that further improvement is possible, particularly in the case of static listening with artificial pinnae, however if the quality of the results generated by Batteau is to be approached, then a substantial increase in the specification for the auditory system will be required due to the specification of microphones, power supplies and pre-amplifiers used. One of the main features of auditory localisation for telepresence is the ability to move the head, which was a feature already present, due to the work of the current Strathclyde team. The advantage of using pinnae is demonstrated but many more trials are required to determine their efficacy, particularly if the pinnae used, are not cast from the user. The addition of vision cues to the system for auditory localisation helps to give an overall sense of "being there" and helps to uniquely specify direction. "Ears tell eyes where to look."[3]

REFERENCES

1 G.M.Mair, J.G.Heng, R.J.Fryer, N.Clark, D.Sheat, An Experiment in Long Distance Telepresence Using ISDN2, Mechatronics 96, Vol. 1, pg. 257-262

2 Robert H Gilkey, The Sense of Presence for the Suddenly Damaged Adult - Implications for Virtual Environments, *Presence*, Vol. 4, No 4, MIT, pg. 357-363.

3 Elizabeth M. Wenzel, Localisation in Virtual Acoustic Displays, *Presence*, Volume 1, Number 1.p 80-107, 1992.

4 William A Yost and Donald W Nielsen, *Fundamentals of Hearing*, 3rd Edition, CBS College Publishing, 1985

5 A. W. Mills, Auditory Localisation, Foundations of Modern Auditory Theory, Vol. 2, Editor J.V. Tobias., Academic Press, 1972.

6 H.W. Gierlich and Klaus Genuit, Processing Artificial-Head Recordings, *Journal of the Audio Engineering Society*, Vol. 37, No1/2 1989.

7 M.D. Burkhard and R.M. Sachs, Anthropometric Manikin for Acoustic Research, *Journal of the Acoustic Society for America*, Vol 58, No1, 1975 p214-221

8 D.G. Malham, Spatial Hearing Mechanisms and Sound Reproduction, University of York, Http://www.york.ac.uk/3d_audio/ambis2.htm, © 1998

9 William M. Hartmann and Andrew Wittenberg, On the Externalization of Auditory Images, Journal of the Acoustical Society of America, Vol. 99, No 6, June1996.

10 H. Wallach, The role of head movements and vestibular and visual cues in sound localisation. *Journal of Experimental Psychology*, Vol. 27, pg. 339-368.

11 D.W. Batteau and R.L. Plante. Localisation of Sound:Part 2. The Mechanism of human localisation of sound with application in remote environments. *U.S. Naval Ordinance Test Station Report, TP 3109 Part 2*. 1962

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A Novel Teleoperated Robot Control System

A. R. Graves and C. A. Czarnecki

Centre for Computational Intelligence, School of Computing Sciences,
De Montfort University, Leicester, UK, LE1 9BH.

Telerobotic systems are commonly used in environments which are considered too hazardous for humans to work in. One such example is the military application of robots for bomb disposal tasks. This paper describes the enhancement of an existing teleoperated bomb disposal vehicle into an advanced mechatronic system. The approach taken is both technology driven and centered around the specific task problems associated with the application area.

1. INTRODUCTION

Remotely disposing of bombs is an important application area in telerobotics. When the degree of risk is considered too high for a human expert to survey or dispose of a suspect device, a teleoperated vehicle is used. The current system described in this paper requires a human operator to control the vehicle by means of a hand-held controller. The operator is in direct and continuous control of the vehicle, receiving feedback from a video camera, and performing all operations at the level of motor and actuator commands.

There is a strong motivation for improving the design of this vehicle, and this paper discusses ways in which this enhancement is being undertaken. An important aspect of this work is a novel control system which reduces the workload on the human operator. The approach grounds the further development of the vehicle both in an understanding of the specific domain problems, and in current technological trends in the field of telerobotics.

2. BACKGROUND

The evolutionary development of bomb disposal robotic vehicles has continued since the 1970s, although the latest production versions still have user interfaces and computational features which would be considered basic from the viewpoint of mainstream computer science.

This section describes a typical vehicle in terms of both the mechanical structure and operator interface. Additionally the problems associated with its use are discussed. These provide the main motivations for improving upon the system's design.

2.1. The Current System

The remote vehicle consists of a 3DOF hydraulically operated boom mounted on a mobile platform. The platform travels on two sets of tracks driven by electric motors. These are run at different speeds to enable steering. The use of a tracked platform allows the robot to traverse rough terrain, and also negotiate obstacles such as stairs. The vehicle also has video cameras and lamps on-board.

Control of the current version is performed via a hand-held control box which allows the operator to manipulate the motors, actuators and lamps. Due to space limitations on the control panel, multiple use is made of some of the switches, with selection being made by a toggle switch. Webb and Robson identify that problems with the use of the vehicle arise from the design of the hand-held control unit [11]. When operating over long distances, the operator controls the robot guided by video images from the camera which only provides a limited perception of the remote environment.

In later versions this electromechanical control box has been replaced by a microcontroller-based unit incorporating a LCD screen. However, control of the vehicle still requires the operator to work with low-level motor and actuator instructions.

Communication between the operator and the vehicle can be accomplished by one of three different transmission media; electrical cable, fibre optic cable, and radio transmission. Each one of these means of communication has a number of advantages and disadvantages associated with it.

The degree of sophistication attainable by the system's user interface was previously limited by the hand-held nature of any operator control boxes. However, current development work incorporates the concept of modern graphical computer displays and user input devices. It is envisaged that future control systems will be mounted inside vehicles from which disposal operations can be conducted. This trend makes possible the application of more sophisticated user interfaces and the introduction of levels of computer assistance for the human operator. For example, Czarnecki describes the replacement of the hand-held electromechanical controller with a microcontroller mounted on the vehicle itself [3]. This emulates the functions of the standard controller and communicates with a PC via an RS422 link, thereby providing the possibility of more advanced user interfaces than is afforded by a hand-held controller.

2.2. Problems with the Existing Configuration

Bomb disposal robots have been employed by the military for many years and during this time a number of practical problems have been identified. Any improvement to the existing design should be targeted towards these real issues. It is proposed that by the combination of sensor technologies, artificial intelligence techniques and an advanced user interface, a number of these problems can be dealt with.

Misinterpretation of the remote environment by the operator resulting from poor quality visual feedback and a general lack of remote spatial awareness can result in the vehicle colliding with obstacles or even becoming trapped in narrow passages. Obstacle collisions with both the carriage and boom must be considered.

The traversal of steep inclines has also been identified as a significant problem. Most mobile robots have only to contend with planar environments. This vehicle is capable of traversing steep inclines, however, it can become unstable and topple over if the weight distribution is incorrect (i.e. by extending the hydraulic boom).

The use of an umbilical cable, when radio communication is not a viable option, can lead to the problem of cable entanglement, either around the vehicle itself or obstacles in the environment. Also fibre optic cable may become damaged should the vehicle run over it.

The final issue is not one which is as readily evident as those previously discussed. However, the reduction of the mental load imposed by the system on the operator is strongly linked to these *concrete* problems. Task components such as avoiding collisions with obstacles and cable entanglement require the operator's attention in addition to activities more directly related to the goal of disposing of the bomb safely.

3. TELEROBOTIC CONTROL SYSTEMS

This section first describes some of the fundamental principles of teleoperation, and how these have been applied to produce a general telerobotic architecture for this application.

There are a number of recognised forms of telerobotic control which differ in the way that responsibilities are divided between the operator and the computational component of the system [8]. These categories may be regarded as existing upon a continuum of autonomy, with the extremes being direct teleoperation and autonomous robotics. This is represented in Fig. 1.

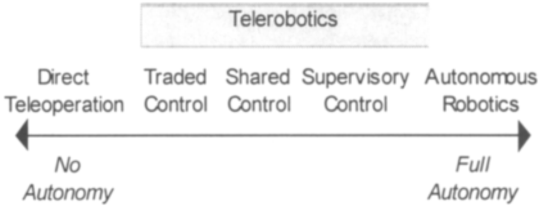


Figure 1. Continuum of telerobotic systems.

In a direct teleoperation system the operator is in complete control and there is no computer present acting as an intermediary between operator and remote device.

In a traded control system a computer is able to control the remote robot and interpret sensor information which the remote device may provide although at any one time either the operator *or* the computer is in control.

A shared control system involves a higher integration of operator and computer competences than is represented in traded control. Typically, any task will consist of a number of sub-tasks, some of which will be performed in parallel. In shared control each of these parallel sub-tasks is allocated to either the operator or the computer. This allows each agent to concentrate upon the sub-tasks which they are most suited to, thus combining the computer's and operator's individual skills in a more productive way than is seen in traded control schemes.

Supervisory control involves the highest degree of autonomy in telerobotic systems. The operator is involved in high level planning, and then instructs the system to complete the task, although the operator supervises and intervenes to solve any problems.

In contrast to teleoperator systems, autonomous robots are those which require no human control, and as such have their behaviours hard-coded in.

3.1. Integration of Control Modes

It must be made clear that the categories of direct, traded, shared and supervisory control are not exclusive modes of operation. It may be desirable for a system to provide a number of different modes of operation which the user may select between. Lin, et. al. [5] suggest that the appropriate control mode is dependant upon the specific type of task being undertaken. Additionally Sheridan [8] states that these control modes may even be in operation simultaneously.

Generally, the degree of autonomy allocated to the system, as defined by the control mode(s), should be used to allow the operator to mediate between the conflicts posed by differing demands and complexity of sub-tasks, their individual critical natures, and the amount of trust placed in the computerised element of the control system.

Due to the safety critical nature the bomb disposal application, it is an important overriding principle that the human operator should always be able to exert full control over the system when required. Although the operator may share many of the tasks with the system, or allocate tasks to it, the operator must always be in a position to trade responsibility for these tasks back should it be necessary.

3.2. Control Loops

Any computational element which constitutes part of a teleoperator system will typically be distributed between the local and remote sites. Automatic control of the system is typically achieved by closing different control loops at both of these two points, as shown in Fig. 2.

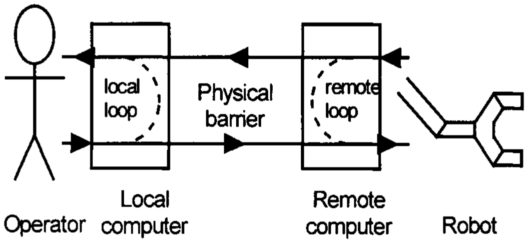


Figure 2. Local and remote control loops.

Control loop closure at the local side takes the form of planning and simulation. At the remote site the computer may interpret sensor information and implement collision avoidance. This is commonly used in situations where teleoperation is time delayed (e.g. teleoperation over the internet, or space telerobotics).

In the bomb disposal system, closure of the remote control loop is used to reduce the amount of data which must be communicated between the two sites. Information from ultrasonic sensors on the vehicle is processed remotely to implement autonomous collision avoidance. Although range information from the sensors can be transmitted back to the local site to provide the operator with a richer description of the environment, this takes the form of pre-processed information, rather than raw data.

3.3. Generic Telerobot Control

The control architecture for this system is shown in Fig. 3. A crucial component of this architecture is the communication protocol which binds the two control systems together.

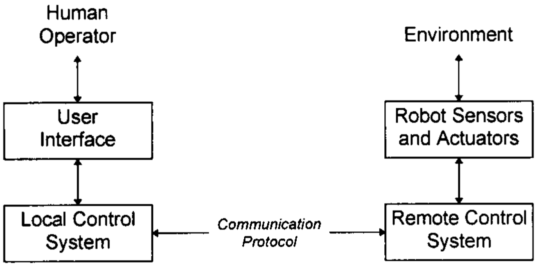


Figure 3. The telerobot control loop.

Given a well-designed user interface it is desirable to make it available to a wide range of telerobots. If the communication protocol can allow the definition and use of services which a telerobot can *supply in an adequately abstract manner*, there is potential for using a single interface to control a variety of different robots. A single consistent interface has the potential to reducing training and equipment costs in application areas where numerous types of robots are required. Military bomb disposal units have access to, among others, standard bomb disposal robots, small surveillance robots and large teleoperated fork-lifts.

To enable the development of generic telerobotic user interfaces, a telerobotic architecture has been designed which is based upon the notion of a number of different types of service which may be provided by a mobile robot. The three types of robotic service defined by this architecture are:

- Core Services - Primitive locomotive actions such as left/right turn and forward.
- Generic Services - Services including the notion of abstract sensor arrays. These services are negotiated during a handshaking process when the system is started up.
- Specific Services - These are services which require the interface to be customised in order for the operator to access them.

The specific services are described in a definition file and ideally, a telerobotic controller will store a version of its own definition file which it can transmit to the local computer upon initialization. This effectively leads to a 'plug-and-play' arrangement whereby individual robots can customize user interfaces to suit their own capabilities. In addition to the representation of physical components such as sensors and actuators, this architecture also includes the notion of interaction with robot *behaviors* which may be traded or shared with the operator.

3.4. Behaviour-based Robotics and Telerobotics

For robots operating in real-time unstructured environments it is accepted that robot control systems based on Brooks' subsumption architecture [2] have a number of distinct advantages. Such systems display a high degree of performance even when running on low-power on-board computers. Also they exhibit the necessary robustness for dealing with real-world environments. Controllers constructed in this way are known as behavior-based systems; their design involves decomposing the high-level problem of robot control into a series of small concurrently running behaviors which tightly couple sensor input to motor output. This structure is in marked contrast to the traditional sense-plan-act cycle and accounts for the high performance achieved by these systems.

Such controllers have been predominantly applied to the field of fully autonomous robots. However, the characteristics of Behaviour-based controllers outlined above make them highly applicable to the field of telerobotics in a number of ways.

The low computational demands make Behaviour-based controllers suitable for implementation on the remote computers on-board mobile robots and provide the necessary robustness for coping with the unstructured or dynamic environments. Additionally, the competences which the Behaviour-based paradigm have been most successful at are those which fit in well with the typical division of responsibility required in telerobotic systems. It has not been possible to build high-level task planning competences out of a substrate of subsumption modules, however, performance at lower levels of behaviour is impressive. Conversely, humans excel at high-level planning activities, even with incomplete or imprecise information, whilst low-level tasks require undemanding, but continuous human attention, needlessly adding to the operator's workload. As with this system, such a combination of human and computer competences is the motivation behind the reactive telerobotic control system described by Arkin [1].

3.5. The Bomb Disposal Telerobotic Control System

Integrating control signals from human operators into Behavior-based controllers poses the designer with a wide range of options. A number of different techniques have been presented in the research literature for accomplishing this. These represent a wide variety of different controller architectures, however, one example based on the subsumption architecture is the space telerobotic system described by Stein and Paul [9]. This controller has the intuitive structure of representing the human operator as the highest level competence in the subsumption network. It is a compelling idea to seek to integrate the human operator into the system as a replacement for the high-level reasoning behaviors which continue to elude subsumption researchers. This general structure is shown in Fig. 4.

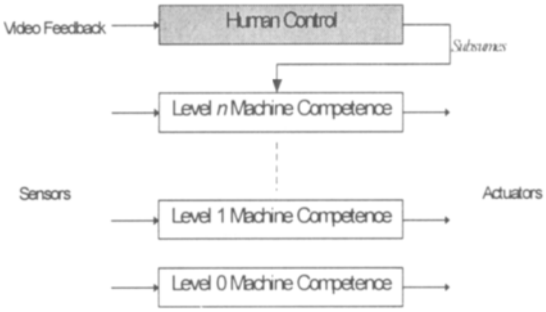

Video Feedback
Human Control
Subsumes
Level *n* Machine Competence
Sensors
Actuators
Level 1 Machine Competence
Level 0 Machine Competence

Figure 4. A subsumption telerobotic controller with the operator as a top layer competence.

However, in telerobotic applications human input is mostly at the level of simple commands (keypresses) and spatial gestures (joystick or mouse movements), rather than high-level communication. It is proposed in this paper that in order to successfully interface the man and machine together, both agents should be represented in the same form. This entails representing the operator as the set of behaviors he/she exhibits, rather than as a single entity. Combined man-machine competences are then achieved by matching corresponding human and system behaviors at the correct level of the subsumption network. Fig. 5 shows the subsumption-based controller for controlling the bomb disposal telerobotic which better reflects this view of the man-machine interface. The human behaviors are distributed around the network rather than grouped together as a single competence layer.

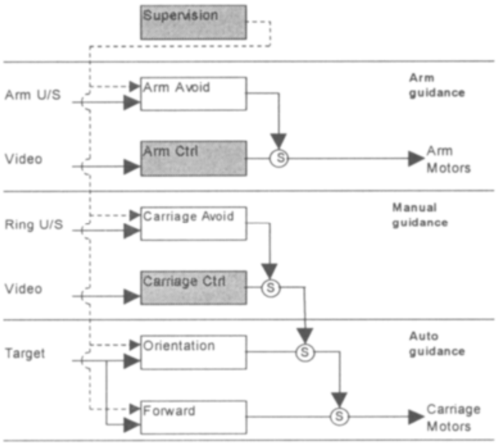


Figure 5. Simplified subsumption network for controlling the bomb disposal telerobot (operator behaviours are shown as shaded boxes).

Interaction between sensor-guided modules and operator modules implements sharing of control in the architecture. The ability of the operator's "Supervision" module to switch on or off individual system modules implements trading of control.

4. TELEOPERATOR USER INTERFACES

Traditional approaches to teleoperator interfaces have reflected a tension between the telepresence and artificial intelligence approaches (see Lumelsky [6]). A more modern, practical approach, described by NASA as *advanced teleoperator systems,* involves a hybrid blend of telepresence technologies, together with graphical overlays to display non-visual data and the use of autonomous functions [12]. This is the approach favoured in this research.

The construction of advanced teleoperator systems involves contending with a large design space of interface options, as the operator must interact not only with the remote environment but with autonomous functions (i.e. behaviors) as well. These characteristics make such systems conceptually closer to traditional GUI application interfaces than pure telepresence systems. Although such systems share many similarities with mainstream user interfaces, they are sufficiently different to mean that existing user interface design and evaluation techniques are not directly applicable. Most existing techniques are geared towards multi-dialogue interfaces and focus upon the ease with which a system can be learnt. Conversely, teleoperator interfaces generally consist of a single information rich screen, and almost without exception users of teleoperator systems undergo substantial training.

There is little useful existing design expertise available for this type of teleoperator interface (Ince, et. al. being a notable exception [4]) therefore development is being guided by the adoption of design principles from other areas, such as Sudano and Marietta's heuristics for complex high-risk systems [10] and Ecological Interface Design [7].

A Window's NT-based user interface is currently being developed as the front-end to the enhanced telerobotic system. This interface includes both standard graphical user interface elements and video feedback from the remote site which is augmented with overlaid video graphics.

5. SUMMARY AND FURTHER WORK

Many of the ideas and concepts outlined in the paper are currently being implemented. The ability to control a number of robots from a single interface means that much of the research work can be performed with a small prototype robot, circumventing the large and cumbersome nature of the actual bomb disposal robot.

The eventual system will represent many of the recent advances in the field of telerobotics whilst applying them usefully to a real world problem. This is being accomplished in a way which recognizes both human factors issues and the technical concerns of integrating sensors and actuators into a distributed control system.

REFERENCES

[1] Arkin, R. C., "Reactive Control as a Substrate for Telerobotic Systems", IEEE AES Systems Magazine, Vol. 6, No. 6, June 1991, pp. 24-31.

[2] Brooks, R. A., "A Robust Layered Control System for a Mobile Robot", IEEE Trans. Robotics and Automation, Vol. 2, No. 1, 1986, pp. 14-23.

[3] Czarnecki, C. A., "Teleoperated Control of Robots", in Proc. EFAM, CCAM, 21-36 Sep. 1997, Ilmenau, Thuringia, Germany, pp. 151-161.

[4] Ince, I., Bryant, K., & Brooks, T., "Virtuality and Reality: A Video/Graphics Environment for Teleoperation", in Proc. IEEE Systems, Man and Cybernetics, 2, pp. 1083-1089, 1991.

[5] Lin, I. S., Wallner, F., & Dillmann, R., "An Advanced Telerobotic Control System for a Mobile Robot With Multi-sensor Feedback", in U. Rembold and R. Dillmann (Eds.), Intelligent Autonomous Systems IAS-4, pp. 365-372, IOS Press, 1995.

[6] Lumelsky, V., "On Human Performance in Telerobotics", IEEE Trans. Systems, Man, and Cybernetics, Vol. 21, No. 5, September 1991, pp. 971-982.

[7] Vicente, K. J., & Rasmussen, J., "Ecological Interface Design: Theoretical Foundations", IEEE Trans. Systems, Man, and Cybernetics, Vol. 22, No. 4, July/August 1992, pp. 589-606.

[8] Sheridan, T.B., "Telerobotics, Automation, and Human Supervisory Control", Cambridge: MIT Press, 1992.

[9] Stein, M.R., & Paul, R.P., "Operator Interaction, for Time-Delayed Teleoperation, with a Behavior-Based Controller", Proc. IEEE Robotics and Automation, Vol. 1, pp. 231-26, San Diego, California, May 8-13, 1994.

[10] Sudano, J.J., & Marietta, M., "Minimizing Human-Machine Interface Failures In High Risk Systems", IEEE AES Systems Magazine, October 1994, pp. 17-20.

[11] Webb, P. F., & Robson, J. I., "A Human Factors Based Approach to The Development of Advanced Teleoperated Robotic Systems", in Proc. Workshop on Recent Advances in Mobile Robots, De Montfort University, UK, pp. 67-71, July 1st, 1997.

[12] Weisbin, C.R., & Montemerlo, M.D., "NASA's Telerobotics Research Program", in Proc. IEEE Robotics and Automation, Vol. 3, pp. 2653-2666, Nice, France, May 12-14, 1992.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Control of the Robot System via Internet

R. Safaric[a], D. W. Calkin[a], R. M. Parkin[a], C.A. Czarnecki[b]

[a]Mechatronics Research Group, Loughborough University,
Loughborough, Leicestershire, LE11 3TU, United Kingdom

[b]School of Computing Sciences, DeMontfort University,
The Gateway, Leicester, Leicestershire, LE1 9BH, United Kingdom

We have launched the Multiple Manipulators for Training and Education (MuMaTe) virtual control and robotics project on the world wide web (www) to evaluate the suitability of virtual learning environments, the internet, and multimedia technologies within an engineering based flexible educational program. A common problem concerns the limited availability of expensive robotics and control equipment with which students in the educational program can work, in order to acquire valuable 'hands on' experience. Students using networked cumputers can access this facility to study a series of interactive experiments with the real world laboratory robot mechanism. This paper presents an experiment which enables students to control a six degree MA2000 robotic manipulator.

1. INTRODUCTION

The increased accessibility to the internet has been successfully exploited by many universities to provide wider access to on-line learning resources. For example, the virtual engineering laboratory developed by Carnegie Mellon University, makes electronic test equipment such as oscilloscopes, function generators, etc available to users across the www, thus introducing students to the concept of remote experimentation [1]. Another example concerns the robot telescope project of the University of Bradford [2], [3]. Other succesful www based robotic projects include the Mercury project [4]. This later evolved in the Telegarden project [5] which used a similiar system of a SCARA manipulator to uncover objects buried within the workspace. Users were able to view the scene as a series of static images and control the position of the robot arm. The university of Western Australia's Telerobot experiment [6] provides internet control of an industrial ASEA IRB-6 robot arm through the www. Users are required to manipulate and stack wooden blocks, and like the-Mercury and Telegarden projects, the view of the workcell is limited to periodically updated static images captured by cameras located around the workspace.

All previously mentioned projects [4]-[6] use cameras to locate and distribute a robot position and environment location to the user via a network. Of course such an approach needs high speed network connections for more or less on-line control of the robot arm. We use a virtual robot arm and environment model instead of cameras minimising the data transmission time through the network so network speed isn't so important as in the mentioned cases. For our robot experiment a 14.4 Kbps modem connection between the user and a real robot system is enough to achieve more or less on-line control of the robot arm.

2. VIRTUAL LAB CONCEPT AND METHODOLOGY

Our virtual laboratory approach is based on the concept that it provides a working facility for hands-on training whilst reducing the need for multiple high cost actual devices. Because the robotic manipulator is only used intermittently the majority of work is undertaken in the virtual environment, where the skills are developed, before final completition on an actual system. Such an approach provides students with greater access to laboratory resources, which is paramount to the education process, but reduces the capital outlay required to provide high quality training environments. It is desirable that the robot simulation should be capable of being executed through any standard WWW browser application e.g. Netscape Navigator etc [7]. But standard browsers for VRML 2.0 language don't incorporate collision detection between shapes in the virtual world [8]. We can solve this problem with building JAVA oriented collision detection software, which is time consuming solution, because JAVA is an interpreted language and it should happen that the virtual world on the user computer wouldn't be executed on-line. Another solution is to build or use finished libraries of complete browser [10] and collision detection software [9] in C-language. We decided to use second approach.

The user has to download the complete MA2000 Robot Simulation application software and installs it first. Communication between the virtual robot model of the robotic manipulator, which is viewed by the remote user, and the control system which positions the joints of the actual laboratory based manipulator is achived as follows:

- the user develops a robot task within the virtual environment,
- transmission of completed robot task file from remote user to the MuMaTE laboratory server,
- authentication, error checking and runtime scheduling of the received task file on the server,
- execution of requested task within laboratory workcell, and
- Collate and return the results to the remote user.

3. VIRTUAL ROBOTICS LABORATORY

The MuMaTe laboratory equipment includes:

- www network server,
- network layer,
- robot workcell, and
- user remote personal computers.

The www network server is responsible for processing the requests for information by an external www browser, installed on the user remote personal computer, delivering on-line documents and providing access to the robotic and control hardware. The server is implemented on an Intel P166 Pentium based personal computer, running the Windows '95 operating system and a www server application program.

The robot workcell utilises the MA2000 six axes robotic arm supplied by Tequipment Ltd. It is designed to be driven from a host computer which then passes configuration and position information for the robots posture to a separate motor control system as each step within the robots programmed task is executed. The motor control system is based around an 8 bit microprocessor (Rockwell 6502) and is responsible for implementating PID servo control for each joint, communication with the host to obtain uptated position set points and controll gains, acquisition of joint positions and current status of peripheral process devices within the workcell. The separate robot control is responsible for a constant sampling time of position control loops and leaving the server free to concentrate on the www interface and the servising of user browser requests via the internet network. So the www server doesn't control the robot mechanism online but instead passes a block of parameters to a dedicated robot controller. The robot workcell is presented in Figure 1.

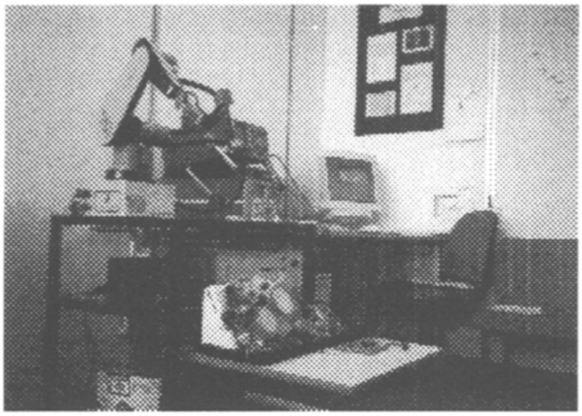


Figure 1: The robot workcell

The network server and local clients are connected to the internet via the campus 10Mbps Ethernet link. Home user clients however utilise a much slower 14.4k or 28.8K modem connection to their local internet service provider using Point to Point Protocol (PPP). A copy of the virtual environment, VRaniML browser's and V-collide's libraries have to be installed on a home clients PCs. This configuration was chosen to allow various interfacing strategies to be investigated, whilst maintaining an open architecture for the future development of the project.

3.1. Software organisation

Following the success of the on-line servo experiment in our Mechatronics laboratory [7], [11] we are now completing the development of an improved human computer interface, integrating C-language and VRML-language within a non immersive desktop virtual reality environment to help improve the realism and sense of presence the user feels when programming the manipulators. This simulation tool will allow the kinematic and dynamic behaviour of the system to be studied, and permit research into task planning, process synchronisation and the communications issues involved with the control of tightly coupled manipulators. A robotic workcell has also been constructed to enable the performance of these novel control paradigms to be studied within a real world environment. The additional processes which are executed by the MuMaTE server and clients remote PC are shown in Figure 2.

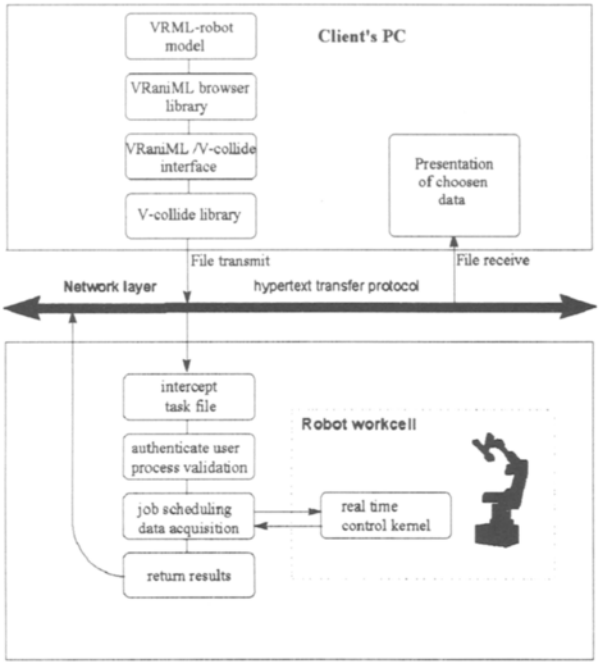

VRML-robot model
Client's PC
VRaniML browser library
VRaniML /V-collide interface
V-collide library
Presentation of choosen data
File transmit
File receive
Network layer
hypertext transfer protocol
intercept task file
Robot workcell
authenticate user process validation
job scheduling data acquisition
real time control kernel
return results

Figure 2: The robot interface between a server and a client

Figure 3 illustrates the users view of the robot model and its assosiated 'virtual' tech pendant.

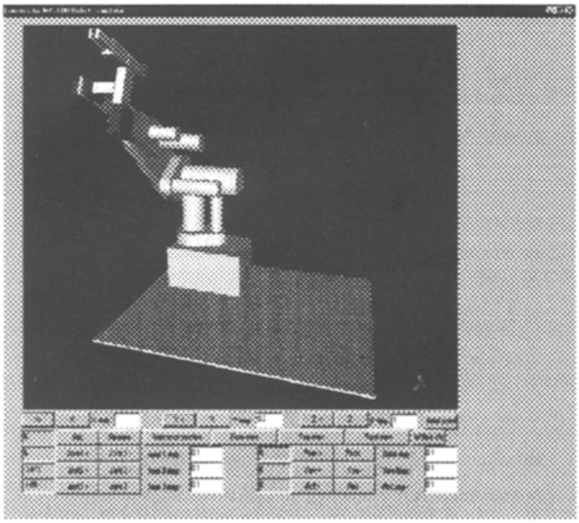


Figure 3: Virtual teach pendant and robot model

3.2. Robot task file transfer

Robot task file tramsmission is run via internet network using hypertext transfer protocol. The ASCII text 'robot task file' (file transmit) adheres to the format shown in figure 4.

```
#program comments

step0:
  5   2   0   0   0    0
500 400 100 500 500 500 0
#this is a park pose for the robot

Step1:
  5   2  0   0   0    0
500 293 78 531 595 500 1
...
#end of file
```

Figure 4: Robot task file

The 'hash' character is used to indetify user comments within the program, whilst 'step0:' etc. inndicates the beginning of a formated block of data representing the next operation to be performed by the robot. The robot's overall task is broken down into a sequential list of intermediate operations or 'steps', each of which is capable of changing the control parameters and pose of the manipulator using the numeric values contained in a data matrix. The default data matrix takes the form presented in table 1.

rate	mode	intput	output	wait		jump
waist	shoul.	elbow	pitch	yaw	roll	grip

Table 1: Default data matrix

The upper row specifies workcell and manipulator control settings whilst the lower row defines joint position data for the current operation. Variable 'rate' determines the speed at which the robot travels when performing the current step. Variable 'mode' alows new function types to be incorporated within the data matrix. Variable 'input' interrogates the status up to four additional peripheral sensing devices within the robot workcell. Variable 'output' actuates any four peripheral output devices connected within the workcell. Variable 'wait' invokes a time delay before proceeding to the next step and value is assumed to be in seconds. Variable 'jump' forces program to jump to a specified step number. Variables 'waist', 'shoulder' and 'elbow' present waist, shoulder and elbow joint position and have values between 000-999 (position between 0 and 270 degrees). Variables 'pitch', 'yaw' and 'roll' present pitch, yaw and roll end effector joint position data and have values between 000-999 (position between 0 and 180 degrees). Variable 'grip' actuates pneumatic gripper (0-closed, 1-open).

3.3. Interface between VRML robot model and VRaniML browser

Figure 5 shows a part of the VRML robot model program written in the VRML 2.0 language. From the code we can see that JOINT2 shape (shoulder) is a child of JOINT1 shape (waist) so JOINT2 shape moves whenever JOINT1 shape moves, but it doesn't happen vice versa.

The VRaniML browser library written in C++ allows user, among others, controlling (rotate) different shapes written in VRML robot model. Figure 6 shows how useful VRaniML browser library is.

The fourth line in figure 6 reads and displays the virtual robot mechanism and a robot environment described in figure 5 on PC screen. Line five finds JOINT2 definition of VRML robot and sets pointer on mJOINT2. Line six is a function which calculates a rotation position of JOINT2, and last line changes the angle in DEF JOINT2 Transform's command in the VRML robot model described in figure 5.

3.4. VRaniML browser and V-collide libraries interface

V-collide is a C++ library for interactive collision detection among arbitrary polygonal mod-

```
...
DEF Robot Transform{
 translation 0.0 0.0 -2.0
 rotation   0.0 0.0 1.0 1.28
 children
  [
  DEF BASE Inline{
   url "basebox.wrl"
  },
  DEF JOINT1 Transform{
   rotation 0.0 0.0 1.0 0.0
   children[
    Inline{
     url "link1.wrl"
    },
   DEF JOINT2 Transform{
    rotation 0.0 1.0 0.0 -0.471
    center 0.0 0.0 260.0
    children[
     Inline{
      url "link2.wrl"
     },
...
]}]}]}
```

Figure 5: VRML robot model program

```
#include "vraniml.h"
...
...
b->ReadURL("c:/robot_c/ma2000s.wrl");
vrTransform *mJOINT2 =
(vrTransform *)b->FindByName("JOINT2");
JOINT2rot=calculate_pos(JOINT2);
mJOINT2->SetRotation(SFRotation
(0, 1, 0, JOINT2rot));
...
```

Figure 6: Rotating JOINT2 with browser

els undergoing rigid motions in VRML environments [9]. This library offers a practical toolkit for performing interactive and robust collision detection in VRML environments [8].

The VRaniML browser library, written in C++, is used for displaying of VRML models and movements of virtual bodies and shapes on the PC screen whilst V-collide library is responsible for preventing collision between virtual bodies and shapes. We had to write an interface between both libraries because there exists significant difference between them how to describe shapes and their rotation and translation.

The VRaniML library uses grammatic rules of the VRML 2.0 language whilst V-collide library uses a homogeneous transformation matrix [12] for description of shapes in the virtual world. For example, VRaniML understands VRML 2.0 description for box (figure 7) with command: geometry Boxsize 2.0 4.0 6.0.

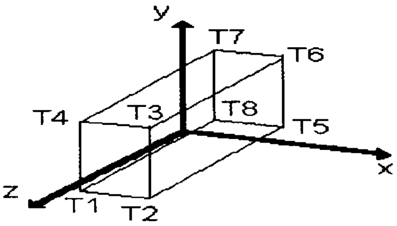


Figure 7: An example

On the other side, we have to describe the same box from figure 7 in V-collide 'language' as series of triangles:

1^{st} triangle: T_1, T_2, T_3
2^{nd} triangle: T_1, T_3, T_4
3^{rd} triangle: T_2, T_5, T_6
4^{th} triangle: T_2, T_6, T_3
...etc

Because such an interpretation in case of complicate shapes in V-collide library would use too much computational time we have made some simplification, so every shape in V-collide 'language' is described by a simple box which is bigger than the original shape.

Interfacing rotation and translation between both libraries is even more complicated because reference coordinate system are not equal. We can see the difference between the coordinate systems in figure 8.

The interface between VRaniML and V-collide libraries is written in table 2.

For example, we can translate VRML robot model program from figure 5:

DEF Robot Transform command can be translated as: $Transl[x, -2].Rot[x, 1.28] =$
$= Transf(Robot)$.

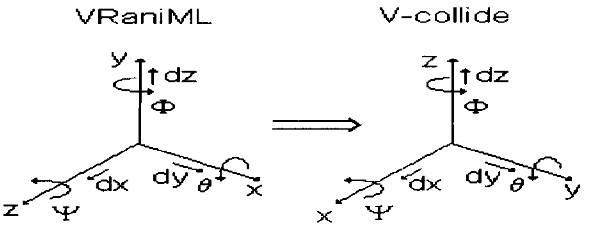


Figure 8: Reference coordinate systems

VRaniML					V-collide
rotation	0.0	0.0	1.0	Ψ	Rot[x,Ψ]
rotation	1.0	0.0	0.0	θ	Rot[y,θ]
rotation	0.0	1.0	0.0	Φ	Rot[z,Φ]
rotation	0.0	dz	0.0		Transl[z,dz]
rotation	dy	0.0	0.0		Transl[y,dy]
rotation	0.0	0.0	dx		Transl[x,dx]

Table 2: Interface between VRaniML and V-collide

DEF JOINT1 Transform command is translated as:
$Transf(Robot).Rot[x, \Psi]$,
where $\Psi = JOINT1rot$.
DEF JOINT2 Transform command is more complicated than previous commands because it rotates the shape JOINT2 about a center point (0.0 0.0 260.0) about y-axis:
$Transf(Robot).Rot[x, \Psi].Transl[x, 260]. .Rot[z, -0.471 + \theta].Transl^{-1}[x, 260]$, where $\theta = JOINT2rot$.

4. CONCLUSIONS

This paper introduced a new www based virtual laboratory for robotics engineering, allowing both interactive simulations and remote experimentation with a real world robot mechanism. Users can configure the robot tip path and parameters, invoke the experiment, and analyse the response of the robot system.

Our distributed control methodology also eliminated the unpredictable network loading problems and variable transmission times faced by any other direct teleoperating systems.

Acknowledgments

This project is founded under the Joint Information Systems Committe's Technology Application Program and the NATO/The Royal Society Program. Robotic and software resources were kindly donated by Mike Millman, TecQuiptment Ltd., Great Hill Corporation [10] and UNC Research Group on Modeling, Physically-Based Simulation and Applications [9].

References

[1] http://www.ece.cmu.edu

[2] J. Baruch, M. Cox, Remote control and robotics: An internet solution, Computing and Control Engineering, Vol. 7.

[3] http://baldrick.eia.brad.ac.uk

[4] K. Goldberg, M. Maschna, S. Gentner, et al., Desktop teleoperation via the www, Proceedings of the IEEE International Conference on Robotics and Automation, pp. 654-659, Japan 1995.

[5] http://telegarten.aec.at

[6] http://telerobot.mech.uwa.edu.au

[7] D. W. Calkin, R.M. Parkin, R. Safaric, C.A. Czarnecki, Visualisation, Simulation and Control Of A Robotic System Using Internet Tecnology, Proceedings of fifth IEEE International Advanced Motion Control Workshop, Coimbra University, Portugal, 1998.

[8] T. C. Hudson, M. C. Lin, J. Cohen, S. Gottschalk, D. Manocha, V-collide: Accelerated Collision Detection for VRML, Proceeding of VRML'97, ACM Press, Monteray, USA, C24-26, 1997, pp. 119-125.

[9] http://www.cs.unc.edu/ geom/V-COLLIDE/

[10] http://greathill.com/download/

[11] D. W. Calkin, R. M. Parkin, C. A. Czarnecki, Providing acces to robot resources via the world wide web, Concurency: Practice and Experience, 1998 (accepted for publication).

[12] K. S. Fu, R. C. Gonzales, C. S. G. Lee, Robotics: Control, Sensing, Vision, and Intelligence, Mc-Graw-Hill Book Company, 1987.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

MAN-MACHINE INTERFACE FOR REMOTE PROGRAMMING AND CONTROL OF NC MACHINE TOOLS AT THE TASK LEVEL: AN EXAMPLE

Francisco Moreira [a], **Goran Putnik** [b], **Rui Sousa** [c]
Production and Systems Engineering Department - School of Engineering
University of Minho, Campus of Azurém, 4800 Guimarães, Portugal
Tel: ++351 (0)53 510278 Fax: ++351 (0)53 510268
[a] Email: fmoreira@eng.uminho.pt [b] Email: putnikgd@eng.uminho.pt [c] Email: rms@eng.uminho.pt

Abstract

This paper presents an example of Man-Machine Interface for remote programming and control of NC machine tools. The system is based on a computer virtual panel which substitutes the machine control panel and in telematics interfaces which establishes the basis for geographical independence between the physical equipment and human operator.

Keywords : Man-Machine Interface, Internet, Teleoperation, CAM.

1. Introduction

The concept of a competitive enterprise today implies, among many others, a capability of flexible access to the best resources. To provide an “ideal” enterprise system, the domain of the selection/employment of resources needed, the “market” of resources, should be as large as possible. We can say, that the competitive enterprise today, and especially in the future, should be capable to manage all manufacturing functions independently of the distance. This is the main idea in developments of the concepts of Global Manufacturing Systems, Distributed Manufacturing Systems or Virtual Enterprise/Factory. This implies the highest level of communication among enterprise’s resources (to be integrated and being integrated), performed in an organised multimedia and information environment in a real-time, that is, using telematics technologies, which can be seen as a key for the future competitiveness. In this paper we present one of our results in development of industrial services based on telematics technologies, especially telematics interfaces for CAD/CAM/CAPP/PPC, which establish, so called, teleservices/teleworking concept.

2. Conventional and Teleoperated CAM System

A conventional milling machine, as an object of teleoperation, has many design problems. Milling machines are designed and made with the assumption that the human operator locally operates the machine control unit panel, Fig. 1a). DNC integrated with CAD/CAM systems do not change the way machines are operated in terms of hypothetical, or possible, (globally) distribution of system components locations. They improve the way part program for machining are built and the way those programs are sent to the machine, Fig. 1b), but the components’ locations are still limited within the local area, which is demonstrated by the use of Local Area (Communication) Network - LAN.

The hierarchical model of control structure for traditional equipment control, based on the model presented in 1979 by Barbera, Albus and Fitzgerald [BARBERA ET AL. 79], is showed in Fig. 1c) (this model is important to understand which control function are subjected to "distribution"). When operating a machine at distance the operator could execute a sequence of "button touches" in a remote and virtual panel, generating, in real time, the expected actions on the machine that is apart from the operation site. The hierarchical control structure for this new system includes some higher levels of control that are apart from the lower levels control functions. We developed a CAM system based on teleoperation at the task level, given in Fig. 2a). This system corresponds to the hierarchical control structure presented in Fig 1d). The informal representation of the logical structure is shown in Fig. 2b) and the corresponding formal representation, in ESTELLE, in Fig. 3. The perception of the state of the machine is also assumed to be recognised by the senses and knowledge of the operator, thus requiring transmission of useful information from machine control unit and external sensors.

System Specification

The formal specification of the system using the Formal Specification Technique (FDT) ESTELLE (Extended Finite State Machine Language)[1] is given bellow. The complete description would be very large, so only a small part is presented here. ESTELLE was chosen because it is particularly adequate to describe distributed or concurrent systems, and models a system as a hierarchical-structured set of communicating non-deterministic state machines (automata). The automata concept is adequate for the specification of production systems. Even the behaviour of the human operator module in this system could be modelled with FDT's (see examples in [Turner 93]). The architecture of the teleoperation system specified by ESTELLE is in fig. 3. Each box is a module and modules communicate through channels with bi-directional capability.

The formal specification is given through channel and module description. The channel description includes the definition of the messages that are allowed to flow in that channel. The channel between Interface1 module and WAN module is defined as:

```
channel Interface1Connection(Interface, Network);
    by Interface:
            MediumInformation(Medium: MIType);
    by Network:
            ReceiveProgram;
            RunProgram;
            Program(file: GProgramType);
```

Messages ReceiveProgram and RunProgram carry no parameters and are only allowed in channel Interface1Connection if originated by the module that plays the role of Network (module WAN in this case). Message MediumInformation carries a parameter Medium of type MIType, which is in fact a record with three fields.

```
type MIType=          {Medium information type}
    record
        Image: ImageType;
        Sound: SoundType;
        Temperature: real;
```

The types ImageType and SoundType are defined in a type declaration as binary files. The type GProgramType in channel description is an ASCII text file containing the "G-program" for machine-tool control unit. Unidirectional channels are obviously allowed and that is the case of the channel inside MachineTool module, between modules MachineActuators and ControlUnit.

```
channel MachineActuatorsServer(Actuators,Control);
    by Control:
            DriversOn;
            DriversOff;
            Xposition(XValue:real);
            Yposition(YValue:real);
            Zposition(ZValue:real);
            Spindle(SValue:real);
            Feed(FValue:real);
```

Specification of a module requires a header definition and a body definition. It is possible to have more than one body to use with the same header. When instantiating the module the appropriated body is chosen. The header definition includes the module's class (process or activity), interaction points, channels involved and, eventually, exported variables.

---

[1] Although especially developed for the field of telecommunications there is nowadays an increasing interest in the use of FDT's in other areas like robotics, security, finance, production systems, software engineering, etc.

```
module ControlUnit systemprocess;
    ip
        A: MachineActuatorsServer(Control);
        S: MachineSensorsServer(Control);
        C: MachineToolServer(Control);
end; {Control unit header}
```

Interaction point A is associated with channel MachineActuatorsServer where the module ControlUnit plays de role of Control.

The body specifies the behaviour of the module, based on finite state machine concept. The transition between states depends on the current state and current input. Usually a transition produces some kind of output, for example an output message through one interaction point of the module.

```
body ControlUnitBody for body;

    state RECEIVING, RUNNING, STOP, ...
    ...
        trans
            ...
            when C.RunProgram
                from RECEIVING to RUNNING
                    begin
                    ...
                        output A.DriversOn;
                    ...
                    end;
            ...

    end; {ControlUnitBody}
```

In this formal specification are defined the data types (G program, image, sound, analog data, binary data) exchanged between modules ***Interface3***, ***MachineTool*** and ***Sensors***. All these data types are used in the definition of ***I3L*** channel.

3. Man-Machine Interface

When machines are to be operated at distance a friendly and robust Man-Machine-Interface (MMI) is required. The problems, emerged from the fact of the person in control being apart from the manufacturing equipment, can be minimised this way.

The remote user of the system must have at each instant the notion of being distant, and, at same time, the illusion of having a real machine panel and a real machine in front of him. This way he will be better prepared to deal with new problems he will face. These new problems arise from communications need in both directions. From machine to remote operator the state of machine: visual and acoustic information, and many other binary and analogue data, like temperature, vibration, machine protection ON/OFF, tool ON/OFF, etc. From remote operator to machine, the commands and programs needed to operate the machine and produce the pieces.

As mentioned by Alan Dix [Dix 94] there is no easy answer to the question "what makes a good interface?". However, Man-Machine Interfaces, like machine control unit panel, have been built for decades. These machine control unit panels, independently of respective builders, have some common characteristics. LED's indicate the status of the machine, rotating buttons for selection of digital values or machine operation mode and other pushing buttons for emergency machine stop, program start, keyboard and screen for program editing and monitoring, etc. The panel distribution of this buttons, the colour and size has a correspondence to their functionality and actuation way.

In the implementation of the remote interface the similarities to the local Man-Machine-Interface are obvious. Almost all the buttons have the same visual appearance, the same disposition in the **virtual control unit panel**, and are operated in the same way, although they are now operated with a mouse click instead of a finger touch or hand rotation. This virtual panel is much like the original machine panel, Fig. 4. The functionality's of the virtual buttons are similar to the physical buttons, but *they hide the conversion to commands generated with button touches in a remote PC monitor and the communications required to send those commands from the remote PC to the machine.*

As consequence the Man-Machine-Interface (local operator-milling machine interface) is substituted by a Human-Computer-Interface, HCI, (remote operator-computer interface) and two Machine-Machine Interfaces (local computer to remote computer interface and local computer to milling machine interface), Fig. 2c).

4. Implementation

A real-time prototype of remote programming and control was implemented on a FANUC 0-M machining centre. This machine is located in the Mechanical Laboratory at University of Minho and the operation room is located in the Laboratory for Automatic Production Systems located in another building of the same complex. Mitsuishi [MITSUISHI et al. 92] describes a similar installation: between a machining centre in Tokyo and a operation room located in the same building, and the same machine controlled from George Washington University. Considering the hierarchical model of the control structure for NC machine tools, the "distribution" of control function is implemented between the level of machine tasks - machine programming (CONTROL LEVEL 4), and the level of interpolation and auxiliary functions (CONTROL LEVEL 3). Of course, for a machine tool case, considering functional requirements of precision and axis' speeds, it is not possible to consider any "distribution" at that level and below, but for the higher levels, the "distribution" of functions, namely system control functions, is welcomed (which is under development).

The development tool used in the construction of the programs was LabVIEW 4.1 for Windows 95. This program development tool uses a graphical programming language – the "G" Language. LabVIEW has the basic graphical elements for building virtual equipment panels. These basic elements are LED's, pressing buttons, rotating buttons, displays, graphs and so on. This elements are combined in a way to form the application Front Panel. All the programming tasks are done in the Block Diagram, which is a pictorial solution to a programming problem. LabVIEW programs are called Virtual Instruments (VI) because their appearance and operation can imitate actual instruments.

The front panel of the program that implements the remote operator-computer interface is shown in the Fig. 4. This program is called the Client VI(**RemoteController**). In the machine site there is another PC that implements the local machine-computer interface. This program is called the Server VI (**Interface1**), and since human does not operate it, it has no front panel. An RS232C communication link connects the machine to the local computer. The Server VI and the Client VI implement, as well, the computer-computer interface, which is working on a Wide Area Network (WAN), and was implemented under TCP/IP protocol, usually known as Internet protocol.

When the milling machine is turned ON and the server application is running in the local PC, Client applications (apart from the machine) can ask for working sessions. To do that a Client establishes a connection to the Server and begins working. The Client application can operate the milling machine in two modes: the manual mode and semi-automatic mode. In manual mode the remote operator makes some "clicks" in the buttons of the virtual machine panel and those "clicks" are converted in commands that will be sent to the local PC, via Internet, and from local PC to milling machine via RS232C. In semi-automatic mode, the remote operator selects a program from a library of programs, located in the remote PC, confirms that the program to be sent is really that one, by watching the code in the virtual machine panel, and "clicks" a button to send and execute the program in the machine.

5. Conclusions

Teleoperation can enable new production paradigms to take place in physical factories. Advanced concepts of production systems, based in global distribution of manufacturing equipment, can establish a new era in the way people do their work in industrial environments. The structure to implement this is based on interfaces, required from the fact that person in control is apart from the equipment location: telematics interfaces and man-machine interfaces are required. A PC based man-machine interface with similarities to the traditional control panel of machines takes advantage of the developed familiarity with buttons types, disposition and functionality, and, at the same time, hides the complexity (in terms of network functions and commands generated through button touches) that his behind the virtual control panel.

Regarding our future work, besides the further developments and optimisation of the system presented, we think two technical issues could be important to mention:

1) The relation between the system we developed and open architectures for NC controllers.

2) New standards for machine interfaces could be considered to enable easy construction of this kind of systems.

The project is being developed at the Centre for Production Systems Engineering (CESP-Centro de Engenharia de Sistemas de Produção) at University of Minho, Portugal, within the Virtual Factories/Distributed Production Systems project (VF/DPS).

References

[BARBERA et al. 79] BARBERA, A.J., ALBUS, J.S., FITZGERALD, M.L., "Hierarchical Control of Robots using Microcomputers ", 9th Industrial Symposium on Industrial Robots, 1979.

[BLATTER and DANNENBERG 92] BLATTER, Meera M., DANNENBERG, Roger B., "Multimedia Interface Design", ACM Press, 1992.

[DIX 94] DIX, Alan, "The Human Interface", Assembly Automation, Vol. 14 no. 3, 1994, pp. 9-13, MCM University Press, 1994.

[MITSUISHI et al. 92] MITSUISHI, Mamoru, NAGAO, Takaaki, HATAMURA, Yotaro, KRAMER, Bruce and WARISAWA, Shin'ichi, "A Manufacturing System for the Global Age", Human Aspects in Computer Integrated Manufacturing, pp. 841-852, Elsevier Science, 1992.

[NI 96a] NI, "LabVIEW User Manual", National Instruments Corporation, 1996.

[NI 96b] NI, "LabVIEW Communications VI Reference Manual", National Instruments Corporation, 1996.

[SHERIDAN 92] SHERIDAN, T.B., "Telerobotics, Automation and Human Supervisory Control", MIT Press, 1992.

[TURNER 93] TURNER, KENNETH J., (ed.), "Using Formal Description Techniques - An Introduction to ESTELLE, LOTOS and SDL", John Wiley & Sons.

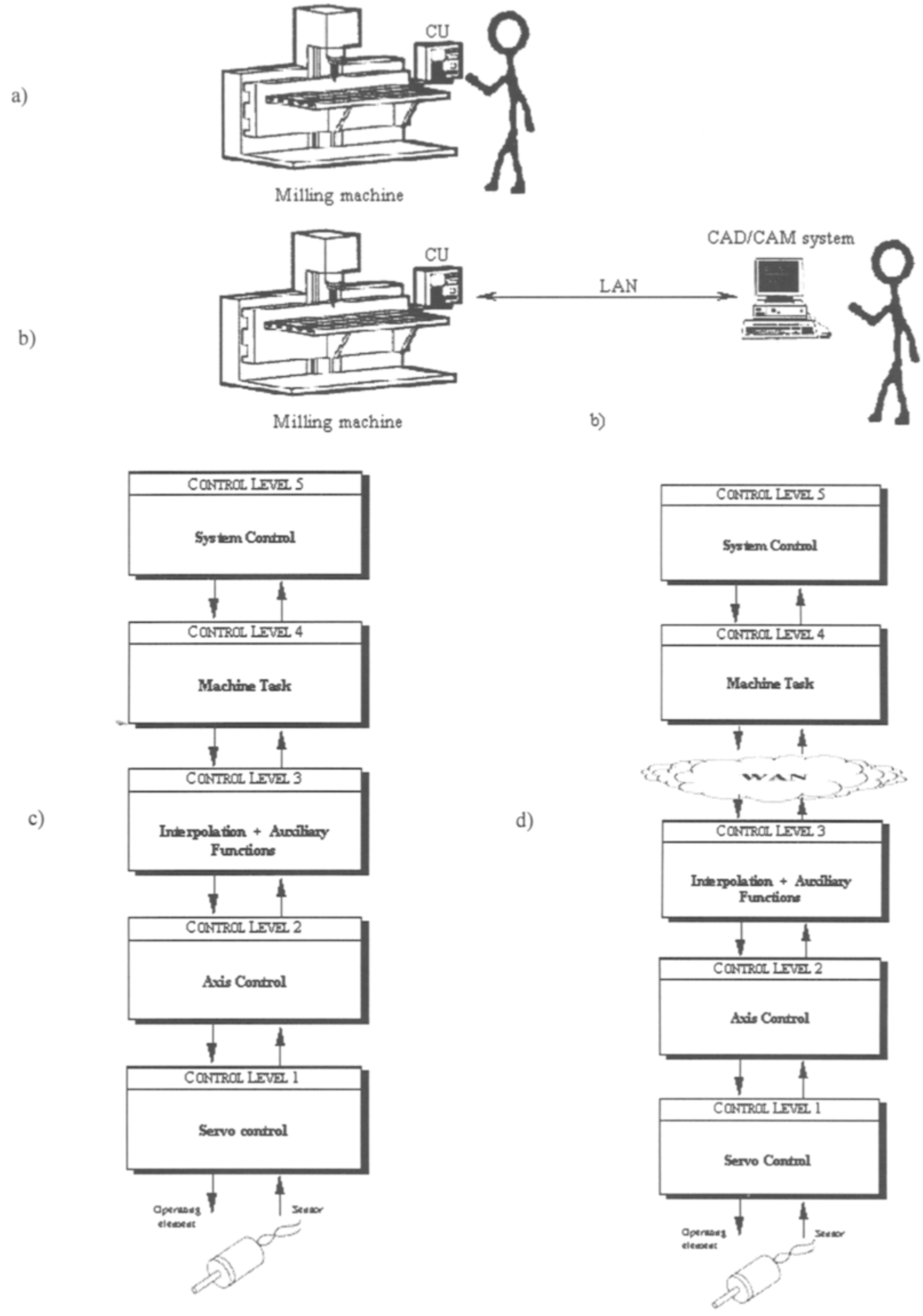

a)
CU
Milling machine
b)
CU
LAN
CAD/CAM system
Milling machine
b)
c)
CONTROL LEVEL 5
System Control
CONTROL LEVEL 4
Machine Task
CONTROL LEVEL 3
Interpolation + Auxiliary Functions
CONTROL LEVEL 2
Axis Control
CONTROL LEVEL 1
Servo control
Operating element
Sensor
d)
CONTROL LEVEL 5
System Control
CONTROL LEVEL 4
Machine Task
WAN
CONTROL LEVEL 3
Interpolation + Auxiliary Functions
CONTROL LEVEL 2
Axis Control
CONTROL LEVEL 1
Servo Control
Operating element
Sensor

Fig. 1 a) CNC system with local control
b) DNC integrated with CAD/CAM system with local area communication network (LAN)
c) Hierarchical control structure of a milling machine
d) Hierarchical control structure of a remote operated milling machine

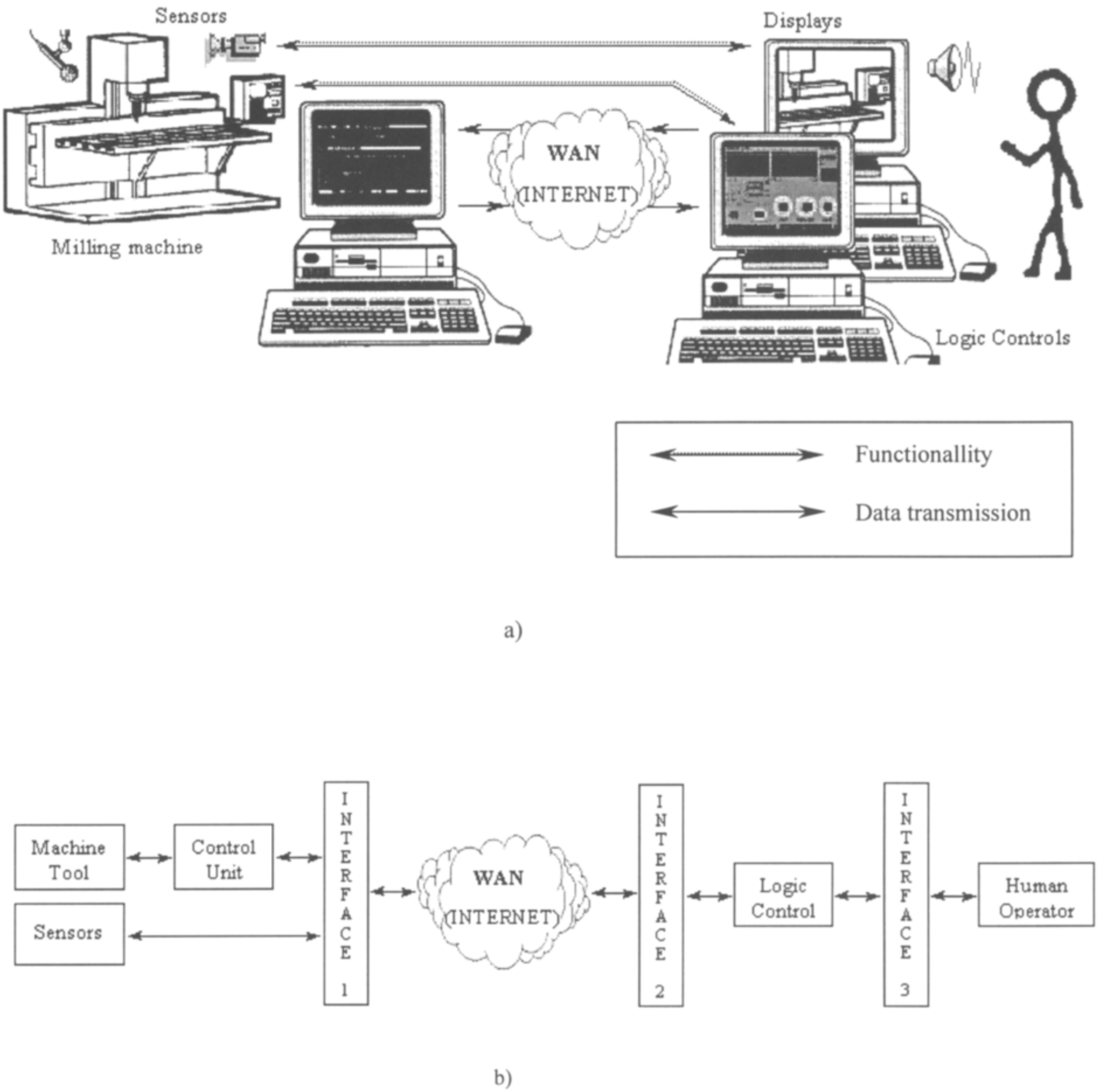

Sensors
Displays
Milling machine
WAN
(INTERNET)
Logic Controls
Functionallity
Data transmission
a)
Machine Tool
Control Unit
Sensors
INTERFACE 1
WAN
(INTERNET)
INTERFACE 2
Logic Control
INTERFACE 3
Human Operator
b)

Fig. 2 a) Remote programming and control of a milling machine
b) Logical struture of interfaces for teleoperation of NC machine tools (unformal representation)

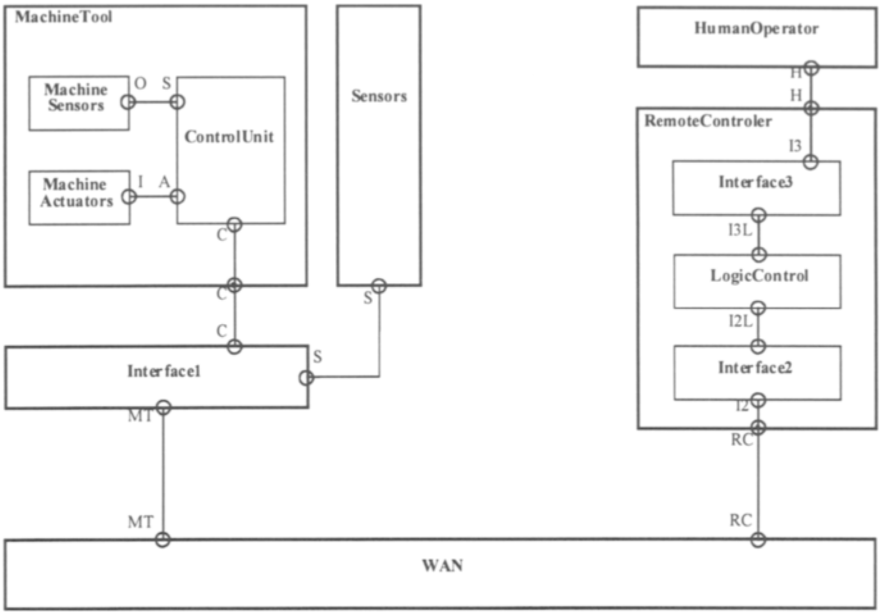
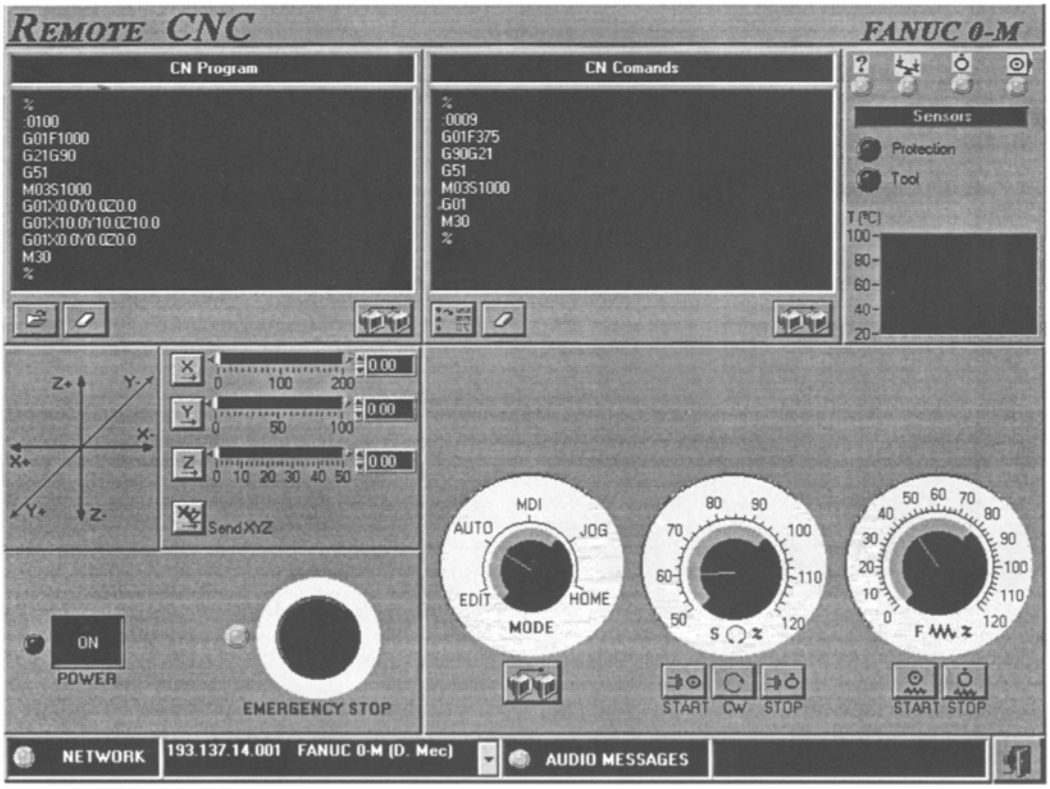

MachineTool
Machine Sensors
O
S
ControlUnit
Machine Actuators
I
A
C
C
C
Sensors
S
Interface1
S
MT
MT
HumanOperator
H
H
RemoteControler
I3
Interface3
I3L
LogicControl
I2L
Interface2
I2
RC
RC
WAN

Fig. 3 ESTELLE formal representation of CAM systems based on teleoperation at the task level.


REMOTE CNC
FANUC 0-M
CN Program
CN Comands
?
Sensors
Protection
Tool
T [°C]
100
80
60
40
20
Z+
Y-
X-
X+
Y+
Z-
X
0
100
200
0.00
Y
0
50
100
0.00
Z
0 10 20 30 40 50
0.00
Send XYZ
ON
POWER
EMERGENCY STOP
MDI
AUTO
JOG
EDIT
HOME
MODE
50
60
70
80
90
100
110
120
S
START
CW
STOP
0
10
20
30
40
50
60
70
80
90
100
110
120
F
START
STOP
NETWORK
193.137.14.001 FANUC 0-M (D. Mec)
AUDIO MESSAGES

```
%
:0100
G01F1000
G21G90
G51
M03S1000
G01X0.0Y0.0Z0.0
G01X10.0Y10.0Z10.0
G01X0.0Y0.0Z0.0
M30
%
```

```
%
:0009
G01F375
G90G21
G51
M03S1000
G01
M30
%
```

Fig. 4 Virtual panel of the FANUC 0-M at the remote site.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Control methods for force reflective master slave systems

H. Flemmer, B. Eriksson and J. Wikander

DAMEK, Department of Machine design, The Royal Institute of Technology
S-100 44 Stockholm, Sweden.

Abstract

This paper presents a force reflecting master slave system under development. A modelling method is presented and with it, some tools for performance and stability analysis. Two methods for force reflection are analysed, implemented, and tested in the specific application where bone milling is the task. Direct force measurement force reflection control and Position error based force reflection control. Direct force measurement force reflection control has the best transparency even though it only supports small force reflection gains with remained stability.

1. Introduction

1.1. Background

The project presented in this paper, Skullbase, is carried out at the Royal Institute of Technology in Stockholm Sweden in cooperation with surgeons from the Karolinska hospital in Stockholm. The project focuses on developing tools for making bone milling in the skull base easier. For certain tumour locations the surgeon has to mill away bone structures in the skull base to reach the cancer tumours lying in the brain tissue. These operations are today performed by the surgeon holding the milling-tool in the hand while bending over the patient in an uncomfortable position. An operation like this is very critical, because of the great risk of causing severe neurological damage on the patient. This is due to the fact that neurons, in contrast to the cells of many other organs, cannot be replenished in an adult individual. Thus inadvertent pressure on the brain or stopping of blood flow in important blood vessels may have devastating effects in terms of loss of neurological function. This “non-forgiving” characteristic of brain surgery has prompted the development of surgical techniques that are “minimally invasive”, meaning that they damage as little as possible. The bone protecting the brain in the skull base is very geometrically complicated, and the surgeon has to maintain high precision in his work through the whole operation. These operations can take up to 20 hours and are very tiring for the team of surgeons working with the patient.

1.2. Approach

A telemanipulated master-slave system prototype has been designed, with intention to give the surgeon a possibility to perform the operation in a more comfortable position and thereby focus all his attention on the milling process. In the telemanipulated system the milling tool is mounted on the slave, which is placed close to the patient

The slave part of the telemanipulator has three active degrees of freedom, two rotational and one translational. The configuration of the links are such that the surgeon has a free view of the total scenery, this is done by using a fairly long last link as seen in the picture of the slave in figure 1.1. By using only three degrees of freedom it was possible to build the slave small in size and keeping it mechanically simple. Backlash is kept low by using Harmonic drive gears.

1.3. Force feedback

It is clear that performing such fine motion and force manipulation require information feedback to the surgeon. Besides audio-visual feedback, force reflection is the main feedback for efficient manipulation.

There were no force data available for milling in bone, therefore the project started with the design of a test equipment to measure these forces. In a single axis rig the milling forces were measured in all three directions for different milling parameters i.e. milling direction, depth, feed velocity, tool rotation speed and tool diameter. Figure 1.2 shows an example measured forces. Forces never exceeded 5 N so the force reflection level is thereby set giving an important design constraint for the master. An other important fact is that the master must be able to reflect very low forces back to the surgeon for fine motion control. Artificial barriers will be implemented to define prohibited areas for the operation giving larger force requirements on the master.

The experimental master is a traditional 2 degree of freedom joystick actuated by two direct drive DC torque motors, se figure 1.3. A design using gears is not appropriate, because of the high motor inertia sensed by the operator as inertia not coming from the slave manipulator. When designing teleoperators it is desirable to aim for high transparency. [1] defines transparency as the extent to which the manipulator copies the feel of the task from the slave to the master. Introducing the concept of impedance from [2] where *force* is equal to *impedance* times *velocity* ($F = Z(s)V(t)$), it is then possible to define a condition for good transparency from [3] as: the impedance felt by the human operator in the master should be a reflection of the environment encountered by the slave.

The rest of this paper will present analysis, stability discussion and some common control methods for bilateral manipulators and some results from tests done with the experimental setup using these control methods. Analysis and testing are performed in one degree of freedom.

2. Analysis and stability

2.1. Analysis

To analyse control schemes it is possible to use the hybrid H-matrix from [4], [5], and [6]. Consider an arbitrary one degree of freedom telemanipulator with force feedback and a master generating references to the slave unit. The one degree of freedom theory can later on be extended to cover the

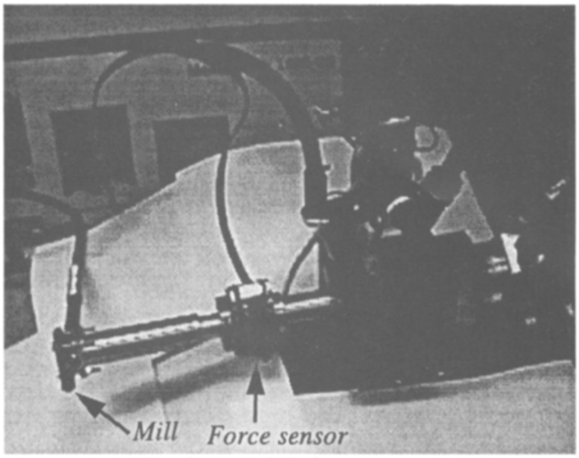


Figure 1.1. The three DOF slave.

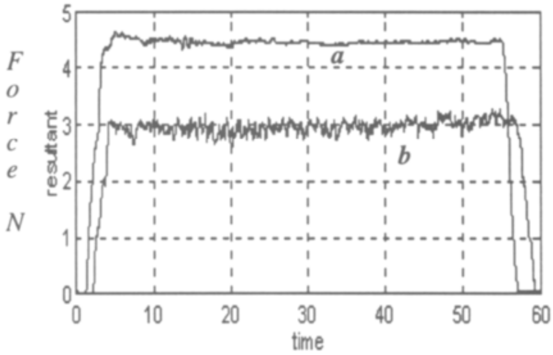


Figure 1.2. Resulting force in Newton acting on the milling tool while milling in bone from a cow. Milling data: tool diameter: 4mm, feed speed: 1 mm/s, rotation speed: 21000 rpm, depth: 1.1 mm (curve a) and 0.85 mm (curve b).

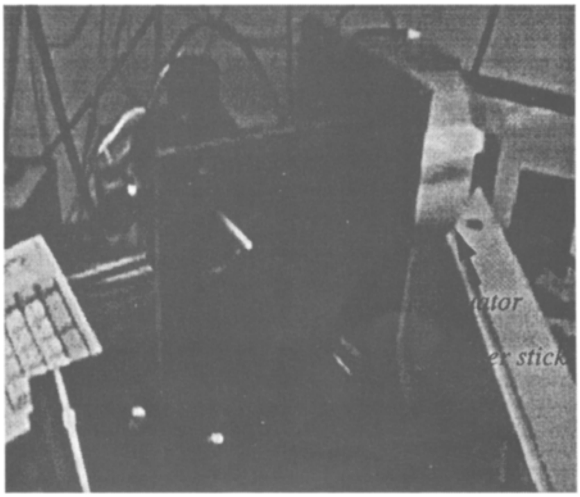


Figure 1.3 The two DOF master stick.

case with n degrees of freedom. The H-matrix relates the independent variables (environmental force F_e and velocity reference $\dot{X}_m$) to the dependent variables (master force F_r and slave velocity $\dot{X}_s$) as:

$$\begin{bmatrix} F_r \\ \dot{X}_s \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \dot{X}_m \\ F_e \end{bmatrix} \qquad (2.1)$$

a mechanical interpretation of the h-parameters is

$$H = \begin{bmatrix} Z_m(s) & G_{fr} \\ G_p & \dfrac{1}{Z_s(s)} \end{bmatrix} \qquad (2.2)$$

where

- G_{fr} is the force reflection gain from contact forces acting at the slave to the master,
- G_p is the position gain from master references to the slave,
- $Z_s(s)$ is the impedance of the slave unit (including servos of the slave).
- F_r is the force in the master stick felt by the operator, and
- $Z_m(s)$ is the impedance in the master unit, force control is passive and open loop.

The ideal teleoperator would have a master unit with a low impedance to obtain high transparency of the teleoperator, a force reflection gain of one and a reference gain (scaling) of one and at last a high output impedance (stiff slave unit) to ensure that the slave maintains desired position. Teleoperator systems may be judged and compared by their H-matrix.

2.2. Stability

To analyse the teleoperator for stability, one idea from [7] and [8] is to derive a transfer function for the teleoperator and study the behaviour of the transfer function for different environments encountered by the slave. Following the ideas in [8] in deriving the transfer function from the operator's desired velocity $\dot{X}_h$ to the true velocity of the master $\dot{X}_m$ as.

$$\dot{X}_m = G_c(s)\dot{X}_h \qquad (2.3)$$

The roots of the denominator of $G_c(s)$ determine the stability of the system. A root locus plot for different environments is a useful tool to judge teleoperator stability. $G_c(s)$ includes models of the human arm, the master, the slave and the environment. Following the ideas from [9] and [8] by modelling the master and the human arm as in figure 2.1.

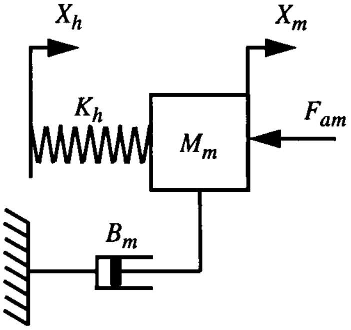


Figure 2.1 Master and human operator. Dynamical model.

[8] models the arm as a spring with stiffness K_h, although the actual arm has a structural damping, the undamped model will provide a worst case in the stability analysis. The master as a mass M_m and a damper B_m. The master is also affected by the actuator force F_{am} from the force reflection. The master and human model is,

$$M_m\ddot{X}_m = -F_{am} - K_h(X_m - X_h) - B_m\dot{X}_m \qquad (2.4)$$

Solving for $\dot{X}_m$ gives,

$$\dot{X}_m\left(M_m s + B_m + \frac{K_h}{s}\right) = \frac{K_h}{s}\dot{X}_h - F_{am} \qquad (2.5)$$

where the impedance $Z_m(s)$ can be identified as

$$Z_m(s) = M_m s + B_m + \frac{K_h}{s} \qquad (2.6)$$

rewriting gives,

$$\frac{K_h}{s}\dot{X}_h = \dot{X}_m Z_m(s) + F_{am} \qquad (2.7)$$

where the right hand side of this expression is defined as the reflected force, giving the force applied by the user.

Continuing with the analysis for the slave, a dynamical model of the slave and its interaction with the environment is needed. Following the ideas of [8] by modelling the slave and the slave controller as a spring with stiffness K_s, a damper with damping B_s and a mass M_s. See fig 2.2.

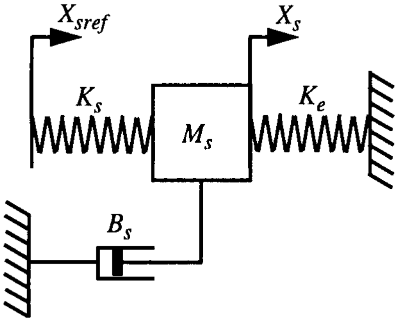


Figure 2.2. Slave and environment, dynamical model.

The damping can be derived from servo damping and joint friction. Slave stiffness comes from the servo controlling the salve. The environment is modelled as a spring with stiffness K_e which may include the stiffness of the tool. Dynamics of the slave is then

$$M_s\ddot{X}_s = -B_s\dot{X}_s + K_s(X_{sref} - X_s) - F_e \tag{2.8}$$

where X_{sref} is the reference position to the servo of the slave. Solving for $\dot{X}_s$ gives

$$\dot{X}_s = \frac{\frac{K_s}{s}}{M_s s + B_s + \frac{K_s}{s}}\dot{X}_{sref} - \frac{F_e}{M_s s + B_s + \frac{K_s}{s}} \tag{2.9}$$

where the denominator can be identified as the impedance of the slave $Z_s(s)$.

3. Control methods

3.1. Direct force measurement force reflection control

One simple form of a force reflecting scheme is called "Direct Force measurement Force Reflection Control" (in [5]) "Forward flow teleoperator") using a force sensor. Slave position references are generated at the master, the force sensor feeds back the sensed forces/torques to the user via the masters actuator. See figure 3.1. Deriving the H-matrix for the control scheme gives:

$$H = \begin{bmatrix} Z_m(s) & G_{fr} \\ \frac{K_sG_p}{sZ_s(s)} & -\frac{1}{Z_s(s)} \end{bmatrix} \tag{3.1}$$

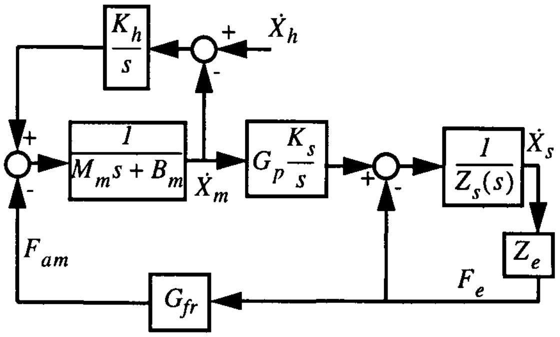


Figure 3.1. Direct force measurement force reflection control

For the stability analysis of the control scheme, the loop gain transfer function $G_c(s)$ 2.3 needs to be derived. The actuator force in the master is $F_{am} = G_{fr}F_e$. Deriving $G_c(s)$ by combining 2.6, 2.7, 2.8, 2.9, and the definition of F_{am} from above gives

$$G_c(s) = \frac{sK_h(Z_s(s) + Z_e(s))}{sZ_m(s)(Z_s(s) + Z_e(s)) + G_{fr}G_pK_sK_e} \tag{3.2}$$

In the root locus plot of $G_c(s)$, (see figure 3.2) the root locations are functions of K_e as K_e increases from 0 (free space motion) to $K_e = 4.5*10^6$.

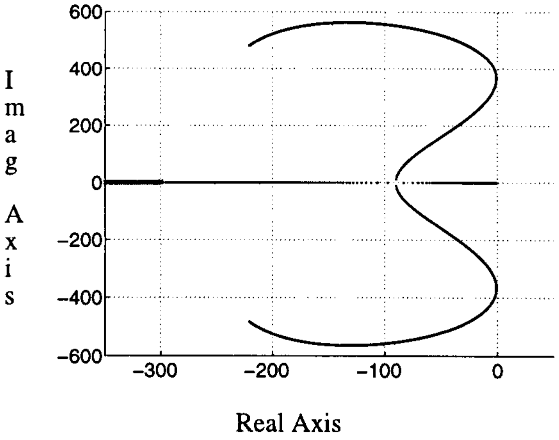


Figure 3.2. Root locus plot of $G_c(s)$ for increasing value of K_e from 0 to 3000, $G_{fr}G_p = 0.15$, $M_s = 1$, $B_s = 600$, $K_s = 90000$, $M_m = 0.05$, $B_m = 22.6$, and $K_h = 11$. Values measured on the experimental setup and from [8].

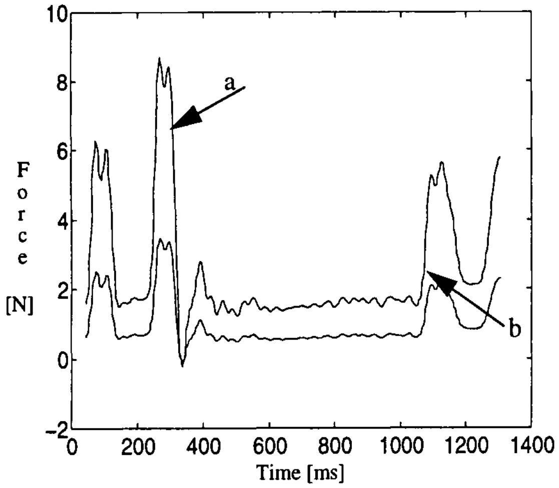


Figure 3.3. Experimental results. Contact force (a), reflected force (b), from direct force measurement force reflection scheme with force reflection gain of 0.375.

The plot is done with a critical value on the product $G_pG_{fr} = 0.15$, a higher value on G_pG_{fr} gives pole locations in the right halfplane for a certain range of K_e. Thus with the actual slave servo and system hardware the force reflection gain is maximized to 0.375 with $G_p = 0.4$. Throughout the measurements with this control scheme G_{fr} was 0.37 which explains the difference between measured force and reflected force in figure 3.3. In Skullbase K_e can vary between different values. When contact occurs K_e will rise quickly to a greater value and then cling off as the mill cuts away bone material, then it is important to have stability for a great range of K_e. The low force reflection gains of this scheme makes it a little insensitive too low forces affecting the slave and thus difficult to use the scheme in a medical application where low forces needs to be sensed by the operator.

3.2. Position error based force reflection control

Another approach to generate force reflection is to let the position error between the master and the slave generate forces to be actuated in the master, namely

$$F_{am} = \frac{G_{fr}}{s}(G_p\dot{X}_m - \dot{X}_s) \qquad (3.3)$$

giving a control structure as in figure 3.4.

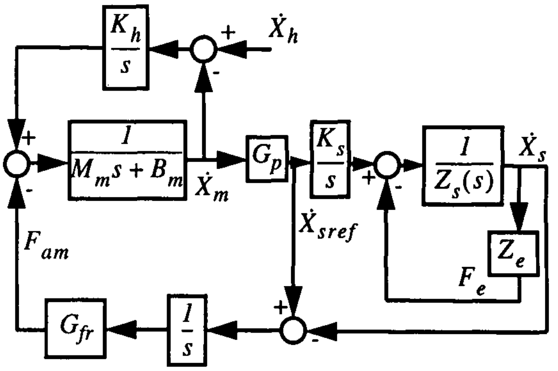


Figure 3.4. Position error based force reflection control scheme.

This scheme is often used in force reflective master slave systems and is often referred to as the "classical" master-slave teleoperator. Our approach differs some from the classical [5] in which both the master actuator force and the slave actuator force is according to equation 3.3. In our approach are position references from the master feeding the position servo of the slave as before, and only the force in the master is according to equation 3.3. Deriving the H-matrix for this scheme gives,

$$H = \begin{bmatrix} Z_m(s) + \frac{G_{fr}G_p}{s}\left(1 - \frac{K_s}{sZ_s(s)}\right) & \frac{G_{fr}}{sZ_s(s)} \\ \frac{K_sG_p}{sZ_s(s)} & -\frac{1}{Z_s(s)} \end{bmatrix} \qquad (3.4)$$

Measurements with this control scheme are performed with $G_{fr} = 0.2$, and $G_p = 0.4$, see figure 3.5. For higher force reflection gains the experimental setup went unstable. The transparency of the control scheme is far from the transparency of an ideal teleoperator. If the slave is moving in free space and the operator moves the master faster than the actual slave motion. A virtual force is sensed in the master during the period of time it takes for the slave to reach the desired position. The operator feels the dynamics of the slave servo. The virtual forces could be desirable if they could be tuned to resemble those of holding the milling tool in the hand while moving it around. One solution to the low transparency is to make the slave more compliant by introducing compliance control from [10], or making the position servo less stiff.

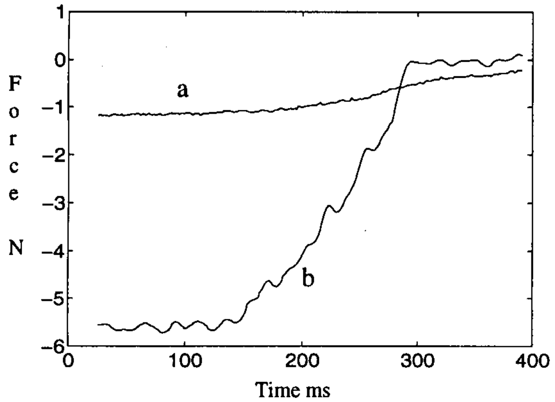


Figure 3.5. Experimental result. (a) Reflected force, (b) measured force from force sensor, with position error based force reflection control scheme. $G_{fr} = 0.2$, and $G_p = 0.4$

4. Discussion

A teleoperator system is designed and built. For characterizing different teleoperator control methods the most common tool is described, the H-matrix. The concept takes into account the two-way signal flow from master to slave unit and from slave to master unit. The model involving the H-matrix is a linearisation of the actual system around some operating point giving the user possibilities to analyse the teleoperator for performance. Performance of the teleoperator can be judged by comparing the H-matrix of the control method with a desired H-matrix. A transfer function for analysis of stäbility of the teleoperator is described based on viewing the teleoperator as a system with a transfer function from desired hand velocity to master velocity. Environmental and human operator impedances are included in the transfer function to get a complete interaction description. Among the different "cases" of operation that can be analysed using the transfer function, one where the slave is encountering an environment with varying stiffness is presented, which might be the case in an application like Skullbase where a mill is milling in bone material. The stiffness of the environment is high in the beginning when contact occurs between the mill and the bone and then decreasing when material is milled away. Two different control schemes are presented, implemented and analysed.

5. Acknowledgements

This project is funded by Branschgruppen för Datorstyrd Mekanik and NUTEK.

References

1. Colgate J. Edward, "Robust impedance shaping telemanipulation", IEEE Transactions on robotics and automation, vol. 9, no 4, August 1993, pp 374-384.
2. Hogan Neville, "Impedance control: An approach to manipulation", Transactions of the ASME Journal of dynamic systems, Measurement, and Control, vol. 107, March 1985, pp 1-24.
3. Lawrence Dale A., "Stability and transparency in bilateral teleoperation", IEEE transactions on robotics and automation, vol. 9, no 5, October 1993, pp 624-637.
4. Anderson Robert J., et. al., "Bilateral control of teleoperators with time delay", IEEE transactions on automatic control, vol. 34, no 5, May 1989, pp 494-501.
5. Hannaford Blake, "A design framework for teleoperators with kinesthetic feedback", IEEE transactions on robotics and automation, vol. 5, no 4, August 1989, pp 426-434.
6. Salcudean Septimiu E., "Control for Teleoperation and Haptic Interfaces", Control problems in robotics and automation, Springer-Verlag, 1998, London, Great Britain, pp 51-66.
7. Hannaford Blake, "Stability and performance tradeoffs in bi-lateral telemanipulation", Proceedings of the IEEE conference on robotics and automation, Scottsdale AZ, vol. 3, pp 1764-1707, May 1989, pp 1764-1767.
8. Bu Yonghong, Daniel R. W., McAree P. R., "Stability analysis of force reflecting telerobotic systems", IEEE Int. Conf. on Intelligent Robotic System, 1996. Vol.3, pp1374-1379.
9. Lee Sukhan, Lee Hahk S., "An advanced teleoperator control system design and evaluation", Proc. of the 1992 IEEE int. conf. on robotics and automation, France May 1992.
10. Kim Won S., "Developments of new force reflecting control schemes and an application to a teleoperation training simulator", Proc. of the 1992 IEEE int. conf. on robotics and automation, Nice, France May 1992.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

ON MODELLING MECHATRONICS SYSTEMS

Bassam A. Hussein, B.Sc., M.Sc.
Production and Quality Engineering Department. The Norwegian University of Science and Technology- NTNU, 7034 Trondheim-Norway
E-mail: Bassam.Hussein@protek.ntnu.no

ABSTRACT

The paper illustrates a systems-oriented approach for modelling mechatronics systems. The physical and logical aspects of mechatronics systems are modeled using multidimensional arrays or generally geometric objects. The consequence of using this array-oriented formalism is that one simulation environment will be sufficient and a few fundamental array operations are needed. The utilization of the procedure is demonstrated for achieving synchronization of multiple independent servomotors.

1. INTRODUCTION

The over all behavior of mechatronics system is determined by the interaction of two basic systems, a continuos physical system and a discrete control system. The continuous system model is the part of the total system that represents the entire continuous-time portion of the mechatronics system. Its state variables evolve with time according to a set of differential equations. Since this part of the system contains the continuous dynamics, it usually contains a conventional continuous time controller (regulator). The control system is a multilevel control structure which is used for implementing various tasks of automatic control that may include; logic control, sequence control, supervision, protection, co-ordination, and etc. [5].

The internal behavior of the control system, i.e. the manner, timing, and sequence in which the control system selects and executes its command signals to the continuos system, is directly related to its logical structure. Hence, the control system could be seen as the container that captures the logical behavior of the overall system. The two systems communicate through an interface that play the role of a translator between the physical and logical variables. Although, much effort has been done in order to analyze this class of systems, yet most of these efforts concentrated on the control configurations rather than common modelling methodology.

What is needed is a **neutral formalism as a basis** for modelling systems that contains both continuos and discrete behavior. Our objective in this paper is to present this neutral formalism for modelling this class of systems. The modelling tool, which we adopt here, is based on utilizing multidimensional arrays as algebraic tool in order to identify the physical and logical behavior of mechatronics systems

A major advantage of this approach is that a small set of operations will apply in general and simulation can be performed using on simulation environment. Simulation of the complete model was carried out by a prototype program for modelling and simulation of multidisciplinary systems called *MSTS-* that was developed at *the Norwegian University for Science and Technology* [1].

2. A SYSTEM MODEL

The system model is divided into two parts, the system and its environment. The mutual influence between the system and the environment are the sources. A system is defined as a set of connected elements. The elements are the building blocks of the system. All relevant properties of the system are contained in those elements. Connections identify how the total system is built up from its basic elements. The approach is application neutral and well suited for modelling mechatronics systems.

3. A SIMPLE SYNCHRONIZATION EXAMPLE

We shall use the following example through out the paper in order to illustrate the scope and implementation of the methodology. In this regard, we must emphasize that we are not concerned with solving typical control problems. We are rather concerned with presenting a framework and an algebraic tool that can be used for modelling and analysis of multidisciplinary systems. Consider for instance a system that consists of two conveyers. Each conveyer is operated by an independent armature-controlled DC-motor. Assume that, at all times the rotating speed of motor$_1$ must be synchronized with the rotating speed of motor$_2$ using the following relation:

$$\omega_1 = 1.5\omega_2 \tag{1}$$

A proportional continuous controller could be installed in each motor in order to force the rotating speed to a certain value so that the synchronization requirements will hold. However, that will remain possible as long as the motors are subjected to pure inertial loads. Occasionally the transfer system may be subjected to non-inertial loads, and consequently, the local proportional controllers would fail to satisfy synchronization requirements. We intend to implement a discrete control system in order to compensate for eventual non-inertial loads so that Equation (1) would be satisfied with minimum error regardless of the operating conditions. Assume that synchronization error (S_{err}) should be kept within the range:

$$(-0.5 < S_{err} < 0.5)\ \text{Rpm}.$$

3.1. Model of the physical system

The differential equations that describe the behavior of the continuous portion of the system will be presented directly because of the limited space available. Complete formulation of the model is given in [3].

$$\begin{bmatrix} e_{f1} \\ e_{a1} \\ e_{z1} \\ e_{f2} \\ e_{a2} \\ e_{z2} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} r_{ds1} & & & & & \\ & r_{qr1} & \psi_{q1} & & & \\ & -\psi_{q1} & r_{z1} & & & \\ & & & r_{ds2} & & \\ & & & & r_{qr2} & \psi_{q2} \\ & & & & -\psi_{q2} & r_{z2} \end{bmatrix} \\ \begin{bmatrix} L_{ds1} & & & & & \\ & L_{qr1} & & & & \\ & & Ir_{z1} & & & \\ & & & L_{ds2} & & \\ & & & & L_{qr2} & \\ & & & & & I_{z2} \end{bmatrix} \end{bmatrix} \begin{bmatrix} q_{f1} \\ q_{a1} \\ q_{z1} \\ q_{f2} \\ q_{a2} \\ q_{z2} \end{bmatrix} \tag{2}$$

Where:

q_f: Electric charge in stator windings.

q_a : Electric charge in rotor windings.

q_z : Angular rotation of the rotor shaft.

e_f : Applied voltage on stator windings.

e_a : Applied voltage on rotor winding.

e_z : Applied torque on the rotor shaft.

r_z: Viscous damping of the rotor shaft

r_{ds} : Electrical resistance of stator windings

r_{dq} : Electrical resistance of rotor windings

ψ_q: Flux density (constant) .

I_z : Moment of inertia including load inertia

L_{ds} : Inductance of the stator windings.

L_{qr} : Inductance of the rotor windings.

The subscripts (1,2) refer to motor$_1$ and motor$_2$ respectively. Layer one in the connected system is zero since no capacitance components exists in the dc-motor model. A unity feedback proportional controller is installed in each motor. The controllers dynamics are given by:

$$e_{a1} = (1500 - \dot{q}_{z1}) \qquad (3)$$

$$e_{a2} = (1000 - \dot{q}_{z2}) \qquad (4)$$

In order to investigate the behavior of the above system without discrete control, simulation was carried under the assumption that motor$_1$ and motor$_2$ were subjected to non-inertial loads that could be given by the following sinusoidal functions:

$$e_{z1} = 2 \times \cos t \qquad (5)$$

$$e_{z2} = \sin 2t \qquad (6)$$

Simulation was carried out by a prototype program for modelling and simulation of multidisciplinary systems (*MSTS*). *MSTS* is based on APL[1] environment. For systems described by differential equations, the simulation is carried out by solving these equations numerically. First, the program transform the differential equations to first order differential equations (state space form) then uses Runge-Kutta procedure in order to solve these equations numerically. Through dialog boxes, the user determines initial conditions and simulation time. Simulation results can then be obtained graphically. Simulation of the above model is shown in Figure 1. The simulation shows that the synchronization error does not lie within the specified range (-0.5,0.5) rpm.

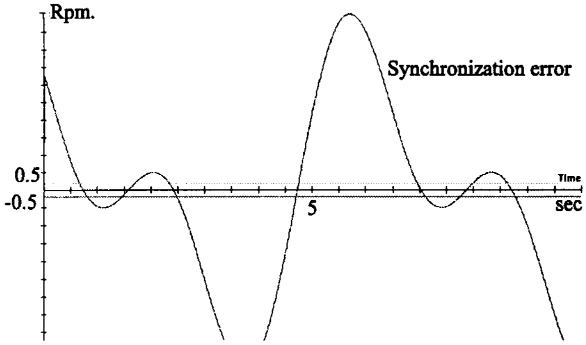


Figure 1. Simulation of the physical system without synchronization control

3.2. Model of the control system

A discrete control system, at any level, represents an abstraction of the physical system or the process to be controlled. Hence, we may consider the control system as a set of connected abstract variables that mirror certain behavior in the physical system. Our aim is to formalize this abstraction as well as the connection between these variables in array theoretic terms. The basic idea is to explore the consequences of considering the truth values as physical measurements. That is the aim is to formalize logic in accordance with the theoretical structure of discrete systems and express this formulization algebraically in array theoretic terms. A major advantage of this approach is that small set of operations will apply in general [2]. In addition, the very same simulation program used for physical systems simulation can be used to simulate the complete system.

---

[1] A Programming Language

In our example, based on measurements of the output speed of both motors, the control system will determine if the synchronization error lies within the The logical behavior of the control system could be described by three logical arguments:

P_1: **IF** S_{11} **THEN** S_{22} **AND** S_{31}

P_2: **IF** S_{12} **THEN** S_{21} **AND** S_{32}

P_3: **IF** S_{13} **THEN** S_{23} **AND** S_{33}

The above arguments are converted into an array form, by performing outer product using APL notation on the binary tubles associated with the States.

$$P_1 \leftarrow \sim(001)\circ.\vee(010)\circ.\wedge(001) \quad (7)$$

specified permissible range or not. If it does not lie within the specified range, the control system will issue some control signals to the physical system to alter the applied armature voltage so that the error will be minimized. In order to meet synchronization requirements, the applied voltage applied on each rotor windings should be forced back or forward by a constant value (*c*).

Based on this description, the discrete control system variables and the associated binary tubles for each state they may undertake as the continuous state space variables evolve in time are in Table 1.

$$P_2 \leftarrow \sim(010)\circ.\vee(001)\circ.\wedge(010) \quad (8)$$

$$P_3 \leftarrow \sim(100)\circ.\vee(100)\circ.\wedge(100) \quad (9)$$

Since all arguments must be jointly satisfied then, the connected system *P* is obtained by performing conjunction on the three arguments and fusing repeated axes:

$$P \leftarrow 1\ 2\ 3\ 1\ 2\ 3\ 1\ 2\ 3\ \text{⍉}\ P_1\circ.\wedge P_2\circ.\wedge P_3) \quad (10)$$

In Array Based Logic, the connected system *P* is the truth table of the control system expressed in multi-dimensional array form as shown in Figure 2.

Control variable	State		
Err Synchronization error	S_{11}= 001 $S_{err} > 0.5$	S_{12} =010 $S_{err} < -0.5$	S_{13} =100 $-0.5 \leq S_{err} \leq 0.5$
M1 Applied voltage on motor$_1$	S_{21}= 001 Increase	S_{22} =010 Decrease	S_{23} =100 Remains unchanged
M2 Applied voltage on motor$_2$	S_{31}= 001 Increase	S_{32} =010 Decrease	S_{33} =100 Remains unchanged

Table 1. Control system variables and their associated binary tubles

Err	s_{11}			s_{12}			s_{13}			
	s_{31}	s_{32}	s_{33}	s_{31}	s_{32}	s_{33}	s_{31}	s_{32}	s_{33}	M2
s_{21}	0	0	0	0	1	0	0	0	0	
s_{22}	1	0	0	0	0	0	0	0	0	
s_{23}	0	0	0	0	0	0	0	0	1	
M1										

Figure 2. Control system with unbounded input.

The connected system *P* expresses all possible states of the entire system after imposing the logical constraints on the structure by connecting its individual elements. The state of each variable the array *P* is obtained by performing reduction on all other axes in the array. The state of all variables must be a tautology in order to ensure that variables are not bounded to any state.

3.3. Interface

The controller and the physical system can not communicate directly, because each utilizes different type of signals. Thus, interface is required in order to convert continuous time signals to controller signals and vice versa. The interface will be consisting of two memoryless mapping functions [6]. One mapping function that converts continuous time signals into bounded inputs for the controller. This function will determine both *when* a continuous signal should bound a control variable and *which* particular controller variable should be bounded.

For the system under consideration, we have only one input variable, which is the synchronization error. This variable will be bounded to one of its three states according to the following function:

$$Err = \begin{cases} S_{11} & if\,(\dot{q}_{z1} - f \times \dot{q}_{z2}) > 0.5 \\ S_{12} & if\,(\dot{q}_{z1} - f \times \dot{q}_{z2}) < -0.5 \\ S_{13} & if\; -0.5 < (\dot{q}_{z1} - f \times \dot{q}_{z2}) \wedge (\dot{q}_{z1} - f \times \dot{q}_{z2}) < 0.5 \end{cases}$$

Assume that the synchronization error was bounded to the state (S_{11}). The resultant multidimensional array in a planer form is shown in Figure 3.

Err	s_{11}			s_{12}			s_{13}			
	s_{31}	s_{32}	s_{33}	s_{31}	s_{32}	s_{33}	s_{31}	s_{32}	s_{33}	M2
s_{21}	0	0	0	0	0	0	0	0	0	
s_{22}	1	0	0	0	0	0	0	0	0	
s_{23}	0	0	0	0	0	0	0	0	0	
M1										

Figure 3. Control system with bounded input.

Bounding the input variable to the state (001) forces the output variables to take new states. These new states can be evaluated by performing reduction on axis 2 and 3 where the output variables are located. Thus, the output variable M1 will be bounded to the state (S_{22} = 010), which means: decrease the applied voltage on motor$_1$, and the output variable M2 will be bounded to the state (S_{31} = 001), which means: increase the applied voltage on motor$_2$.

Interpreting the controller output and converting them to physical signals is performed by the second mapping function. The input to the physical system can take a constant value where each value is associated with a certain controller output. For example, mapping functions for motor$_1$will be given by:

$$e_{a1} = e_{a1} + \begin{cases} c & if\ \mathrm{M1} = S_{21} \\ -c & if\ \mathrm{M1} = S_{22} \\ 0 & if\ \mathrm{M1} = S_{23} \end{cases}$$

Thus, the actual interaction between those two systems will take place externally, through their environment by means of the impressed sources on both systems as shown in Figure 4.

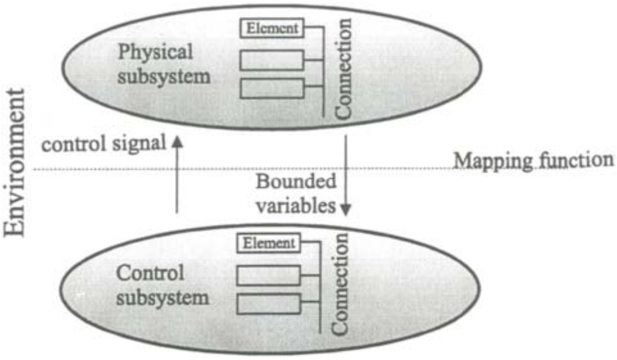


Figure 4. Interface

The physical system generates continuos physical variables. These variables are mirrored by the mapping function to the control system, consequently, some logical variables will be bounded to a certain state and the control system will consequently attain a new state. The new state, which represents the state of the external constraints joined with the system constraints, will generate a new set of controller output. The second mapping function will interpret the controller output and initiate a new set of physical signals to the actuators that derive the physical system. The two mapping functions are added to the program in order to generate the applied sources at each simulation increment. The result is shown in Figure 5.

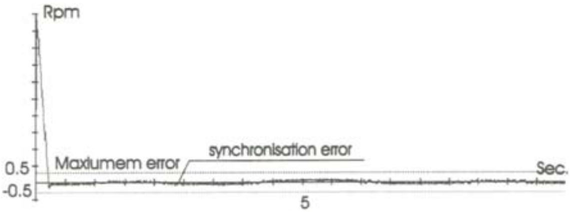


Figure 5. Simulation with discrete event controller.

The simulation shows that the discrete control system kept the synchronization error within the specified range. The system has also been tested for other loading with similar results.

4. CONCLUSIONS

What we have presented here is an array based modelling procedure. The procedure is applied for modelling the physical and logical behavior of mechatronics system using the same type of formulation. In this paper, we have shown that the interaction between these systems takes place externally through the imposed sources on each system. As the physical state variables evolve with time, some logical variables will be bounded to a certain state and the output variables will attain a new state. This new state represents the response of the control system to the external constraints joined with the internal constraints. The bounded output variables will initiate a new set of control signals to be applied on the physical system. The above approach demonstrated its ability to handle the multidisciplinary nature of mechatronics system in achieving synchronization of two independent servomotors.

REFERENCES

1. Øyvind Bjørke, Manufacturing systems theory - A geometric approach to connection. Tapir-Trondheim, 1995.
2. Ole, I. Franksen, The geometry of logic, from truth tables to nested arrays. The 4th international symposium on systems analysis and simulations. Berlin, 1992.
3. Kron, Gabriel, Non-Riemanian dynamics of rotating electrical machinery. Journal of mathematics and physics, 1934, pp.103-194
4. Gert L. Møller, On the technology of array based logic. Ph.D. dissertation Electric power engineering department, Technical University of Denmark, 1995.
5. Rolf Isermann, Digital control systems, Spring Verlag, 1989
6. Antsaklis, et. al. 1993, Hybrid system modelling and autonomous control systems. Hybrid systems workshop, Technical university of Denmark.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Visual-MOOMo –
A Graphical Object-Oriented Modelling Environment for the Design of Mechatronic Systems

Matthias Wolf

Mechatronics Laboratory Paderborn, Prof. Dr.-Ing. Joachim Lückel
Universität-GH Paderborn, Pohlweg 55, D – 33098 Paderborn
Phone: (++49) 5251 60-2421, Fax: (++49) 5251 60-3550
E-Mail: matthias.wolf@mlap.uni-paderborn.de
Internet: http://mlap.uni-paderborn.de

In the design of mechatronic systems, a determined range of functions will be obtained by combining components from mechanics, electrical engineering, hydraulics, and information processing. This combination of highly different disciplines brings about very complex and homogeneous systems to be mastered only by experienced developers.

The present paper describes an interactive tool supporting the developer in his designing mechatronic systems not only on the mathematical level, but also on the physical-topological one. It comprises also a concept for managing complex system models.

1 MOTIVATION

When designing innovative mechatronic products, e.g., an active suspension for a passenger car, the developer builds up a prototype at a very early stage and do comprehensive studies of the real system in order to obtain information on the dynamic system behaviour. Yet it is far less costly and much safer to generate a model of the real system in the computer to determine the dynamic behaviour early on. As mechatronic systems are marked by a combination of components from mechanics, electrical

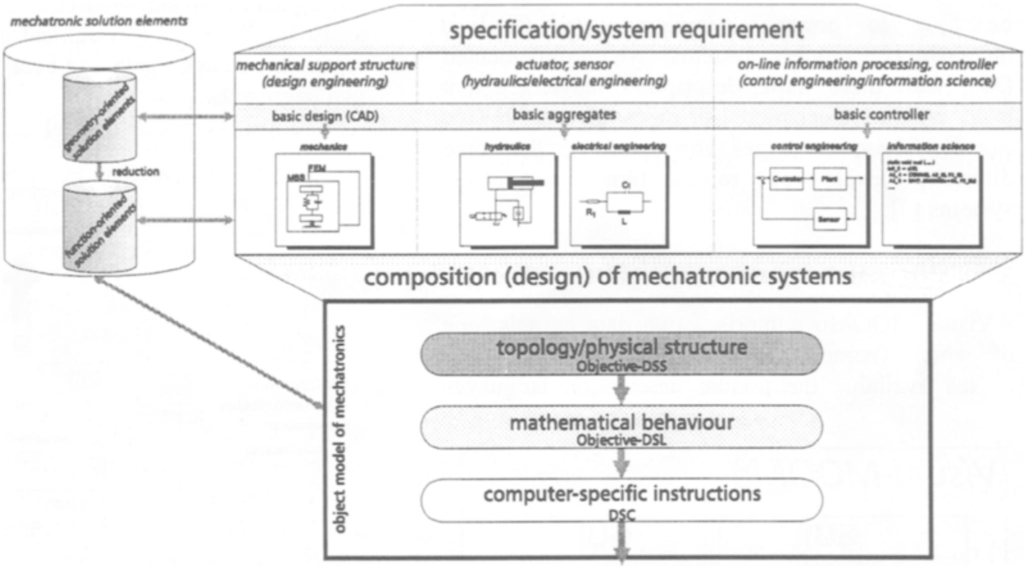

mechatronic solution elements
geometry-oriented solution elements
reduction
function-oriented solution elements
specification/system requirement
mechanical support structure (design engineering)
actuator, sensor (hydraulics/electrical engineering)
on-line information processing, controller (control engineering/information science)
basic design (CAD)
basic aggregates
basic controller
mechanics
hydraulics
electrical engineering
control engineering
information science
composition (design) of mechatronic systems
object model of mechatronics
topology/physical structure
Objective-DSS
mathematical behaviour
Objective-DSL
computer-specific instructions
DSC

Fig. 1- Object model of mechatronics

engineering, hydraulics, and information processing, it is usually very difficult to develop a model.

To do analyses and syntheses on the basis of the mathematical model, the developer must find easy and comfortable means to describe the topological structure and the mathematical behaviour of a mechatronic system. Moreover, the knowledge and experience reflected in this model have to be managed in a way as to guarantee a high degree of safety and resuability.

2 INTRODUCTION

Mechatronic systems are often marked by complex and/or dynamic motions whose representation by mathematical description is becoming ever more intricate. Past experiences with the formulation of models in high-level languages have induced MLaP in the Eighties to develop the model description language DSL (**D**ynamic **S**ystem **L**anguage) [1] that allows to describe mechatronic systems in a textual way on the basis of the scalar, explicit state-space representation form. With small and medium-sized systems this language has proved a success, yet with more complex systems, e.g., those that occur in the modelling of multibody systems, it will become inefficient because it allows only scalar values. To be able to organize, compute, and analyze increasingly complex systems, we complemented DSL with three more description languages that make up the object model of mechatronics [2] (see Fig. 1). These languages correspond to the three different levels of representing mechatronic systems [3].

3 THE CONCEPT OF VISUAL-MOOMo

Visual-MOOMo comprises two data kernels, one of which (named "MOOMo" in the following) makes available the textual description languages Objective-DSS (**Objective**-**D**ynamic **S**ystem **S**tructure) and Objective-DSL (**Objective**-**D**ynamic **S**ystem **L**anguage) that were designed at MLaP for an object-oriented description of modular-hierarchical systems [4]. It offers access to the data-systems-related description language DSC (**D**ynamic **S**ystem **C**ode) that supplies computer-specific instructions for efficient linearization, simulation, and parallelization [5]. The other data kernel takes charge of model management, called "database" in the following (see Fig. 2).

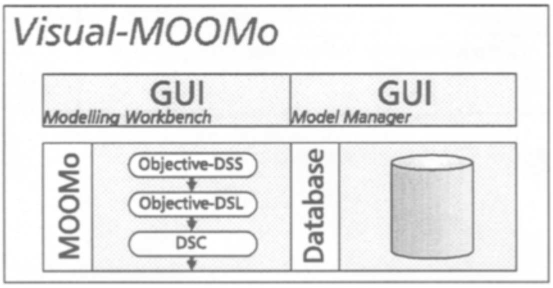


Fig. 2 - Concept of Visual-MOOMo

3.1 MOOMo

MOOMo comprises a number of discipline-specific description classes to represent systems related to mechanics, control engineering, and hydraulics that are available to the user in the shape of classes. These classes and the objects derived from them allow the user to describe complex mechatronic systems. In order to represent the different mechatronic systems in software, we at MLaP developed and implemented an abstract data structure; in the following, it will be referred to as "hierarchical graph" [6]. This development is based on directed graphs and offers many algorithms (BFS, DFS, spanning trees, etc.) for structural analysis. Within a hierarchical graph, every vertex represents an aggregate and every arc, a coupling of two aggregates.

Moreover, this data kernel comprises derivation formalisms (method of force coupling, recursive

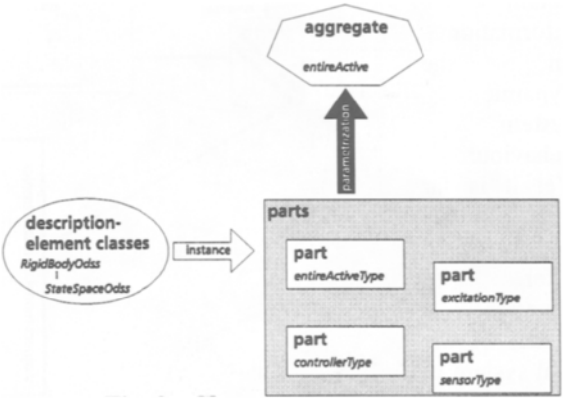


Fig. 3 - Class - part - aggregate

method, minimal coordinates) that derive a mathematical model from the physical-topological one. On the basis of the former model, an analysis (nonlinear or linear simulation) and a synthesis (nonlinear and linear optimization) can be performed. Furthermore, it serves to generate computer-specific instructions that you can use e.g. for parallelization on a real-time hardware.

The discipline-specific classes allow the user to describe parts by indicating the description-element class, followed by a name. Such a part (instance of a class) can be marked by means of the Objective-DSS language definition, i.e., the user defines whether he will have a wheel body or a control arm (see Fig. 3). To produce an actual aggregate, one can supply a part with parameters that will be discriminated in the modelling of a mechatronic system (parameterization of a part).

3.2 Database

In order to support the user in the management and maintenance of these parts and aggregates, we developed a concept that allows to build up hierarchical model libraries (controllers, actuators, sensors, etc.) that are reusable.

Every library, consisting either of references to discipline-specific description classes or of a number of parts and aggregates, will be stored in a repository, called "folder“ in the following (see Fig. 4). Those folders that yield references to the description classes are loaded as default features; all other folders can be loaded at will. Thus every developer can configure his own modelling environment.

When reloading, the user may choose between two formats: one is a binary format in which all parts are saved as objects. This allows quicker loading of the folders, yet a change in a discipline-specific description-element class, due to a version update, may cause problems. In the other format, the parts are described in textual form which will then be read in; in this way, the parts (as objects of a class) are generated. When being read in, the textual variation takes longer to load, but at any time the parts can be reconstructed.

In the modelling process, the user has at his disposal a multitude of parts from the different folders. He has the opportunity to use the part as reference or make a copy. When a reference is made, the part will exist only once. This offers the advantage of requiring less storage location and of simplifying maintenance of the parts, because a change in e.g. part XYZ will automatically effect all parts XYZ. This holds also for the case of one part being used several times over within one model. It is therefore unnecessary to adapt all parts manually in case an error was detected in one part.

4 MODELLING TOOL

In the modelling of mechatronic systems the user is supported by a graphical user interface, named modelling workbench, that allows a quick and easy graphical compilation of systems, with not only the modular-hierarchical structure, but also the connective structure of the systems being visualized [7]. The modelling workbench is conceived in a way as to display only one hierarchical level at a time, thus avoiding displaying too many windows on each hierarchical level. Yet, by using the mouse, one can easily navigate through the different levels of the model. When navigating through the model, the user has access to the description of the aggregates, the input/output relations, and the physical couplings, e.g., joints or hydraulic transmissions. For the individual aggregates, there are different

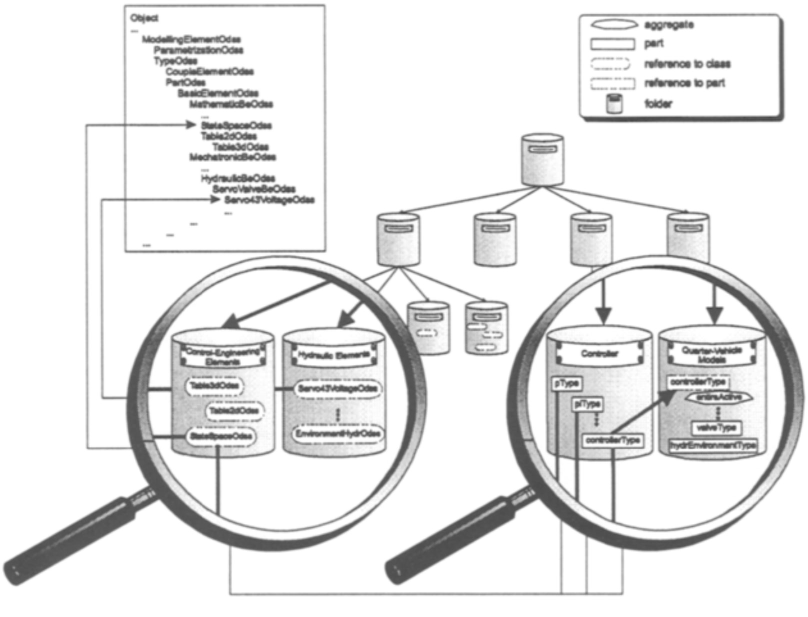

Object
ModellingElementOdes
ParametrizationOdes
TypeOdes
CoupleElementOdes
PartOdes
BasicElementOdes
MathematicBeOdes
StateSpaceOdes
Table2dOdes
Table3dOdes
MechatronicBeOdes
HydraulicBeOdes
ServoValveBeOdes
Servo43VoltageOdes
aggregate
part
reference to class
reference to part
folder
Control-Engineering Elements
Table3dOdes
Table2dOdes
StateSpaceOdes
Hydraulic Elements
Servo43VoltageOdes
EnvironmentHydrOdes
Controller
pType
piType
controllerType
Quarter-Vehicle Models
controllerType
entireActive
valveType
hydrEnvironmentType

Fig. 4 – Model management

representations (*icons*) symbolizing the particular features of the aggregates. Thus the user gains quick access to the models, parts and aggregates being put at his disposal in the folders. Moreover, the physical couplings (e.g., joints) are also marked individually so as to reveal at once which aggregates have been interlinked. If a model is altered the procedure will be checked for admissibility; in this way the model will always be consistent.

As the parts and aggregates always exist in textual form and contain no information on their graphical shape, the hierarchical graph is used to derive the graphical information, e.g., on the position of the blocks or couplings, by means of a modified *placing* and *routing* algorithm (see below) [8]. As the model and the graphical representation are decoupled, changes in the data kernel of MOOMo have no real impact on the graphics. Furthermore, it is possible to employ MOOMo and the database even without the graphical part, i.e., by means of scripts describing for instance a process run.

5 CONCEPTIONAL ASPECTS OF IMPLEMENTATION

In order to point out the lines that represent the direction of the information flow, this chapter will deal with two concepts that are important for a visualization of those models which have no graphical information.

5.1 Placing

Placing requires the hierarchical graph and the area on which the graphical objects are to be displayed.

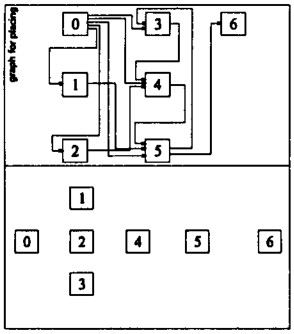


Fig. 5 - Placing

The hierarchical graph serves to calculate the position of the graphical objects representing the aggregates. This procedure is realized as follows: as only one hierarchical level is displayed at a time, it can be determined in the graph which and how many aggregates are to be found on the hierarchical level to be visualized. Now, out of the assembly of aggregates the one is selected that has the most outputs. This aggregate will be placed at the far left. Then, in the hierarchical graph, the placing algorithm iterates over the aggregates of the hierarchical level, by means of a breadth-first search. The breadth-first search offers the advantage of "undulating" over the individual parts which will then be projected in columns on the surface/interface (see Fig. 5).

5.2 Routing

After the positions of the individual parts have been determined, we will find out in the hierarchical graph which and how many connections are located on the hierarchical level in view. Every connection has information on which coupling points of the aggregates are interlinked. This informations and the predetermined graphical object yield the position of the coupling points. One obtains thus two points (source point and target point) that can be linked by means of a router. For this purpose, the routing algorithm that is employed in the VLSI (very large scale integration) to wire channels has been modified.

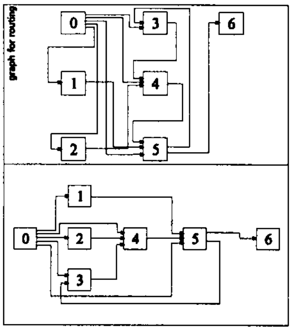


Fig. 6 - Routing

As the wiring of channels on a printed board will mostly be computed just once (this means that time does not matter), the algorithm had to be altered in a way as to become as efficient and fast as possible. The placing as described above was the first step in optimizing the routing because a presorting of the aggregates brings about shorter connections. Moreover, we spread a grid over the plane on which the connections were to be determined. This grid, which is dynamical, reduces the number of points to be visited. Dynamical means that where obstacles occur, the grid was chosen to be very fine-meshed whereas on empty space it can be wide (see Fig. 6).

6 EXAMPLE

The opportunities of Visual-MOOMo will now be exemplified by a wheel suspension with an active strut. The entire physical-topological model consists of an excitation (*excitationC*), evaluation (*evaluationC*), quarter vehicle (*quarterVehicle*), and a controller (*controllerC*) (see Fig. 7). On the left-hand side of the modelling workbench you see the topological structure of the model, by means of which you can navigate through the levels of the model. If you move the mouse across a connection or a coupling point, information will be displayed in the lower information panel, e.g., the name, the class, etc..

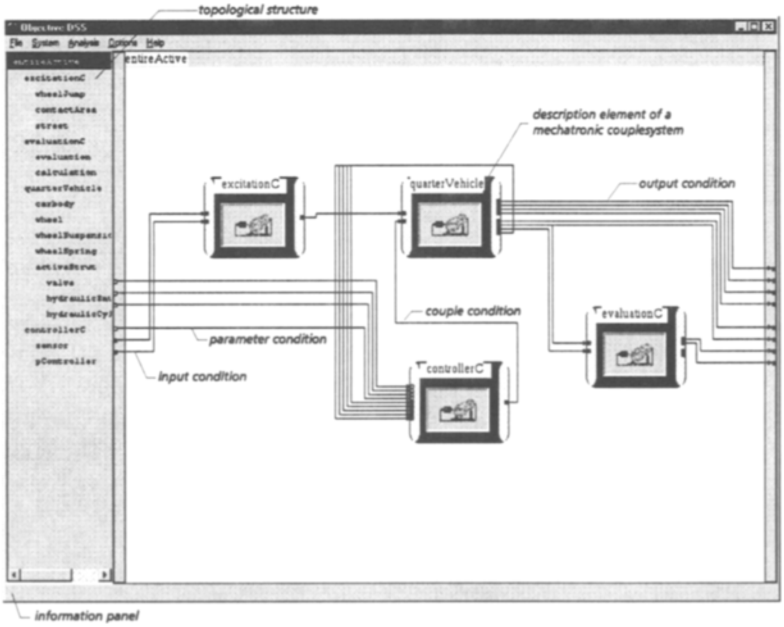

topological structure
Objective DSS
File System Analysis Options Help
entireActive
excitationC
wheelJump
contactArea
street
evaluationC
evaluation
calculation
quarterVehicle
carbody
wheel
wheelSuspensi
wheelSpring
activeStrut
valve
hydraulicEn
hydraulicCy
controllerC
sensor
pController
entireActive
description element of a mechatronic couplesystem
excitationC
quarterVehicle
output condition
couple condition
evaluationC
parameter condition
controllerC
input condition
information panel

Fig. 7 - Modelling workbench *entireActive*

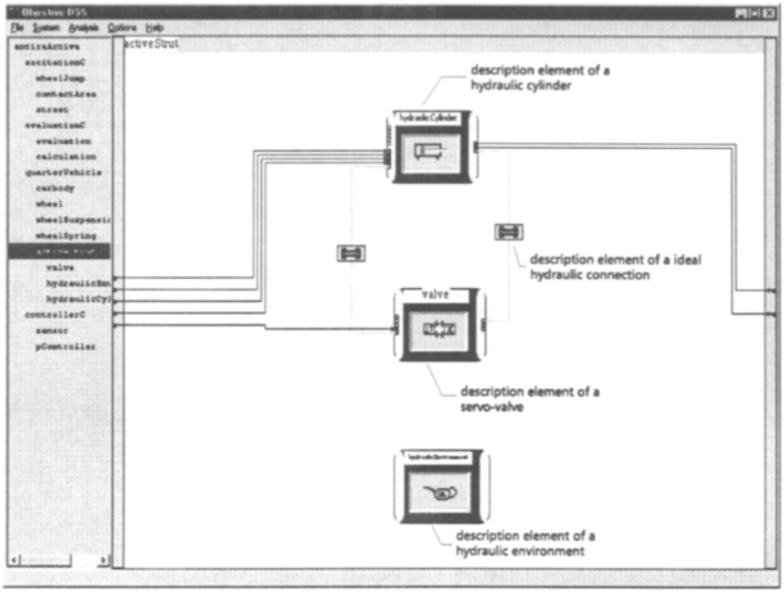
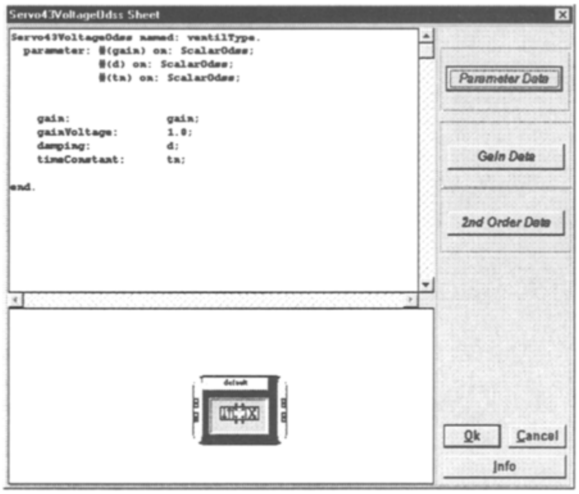

Objective DSS
File System Analysis Options Help
entireActive
excitationC
wheelJump
contactArea
street
evaluationC
evaluation
calculation
quarterVehicle
carbody
wheel
wheelSuspensi
wheelSpring
valve
hydraulicEn
hydraulicCy
controllerC
sensor
pController
activeStrut
description element of a hydraulic cylinder
hydraulicCylinder
description element of a ideal hydraulic connection
valve
description element of a servo-valve
description element of a hydraulic environment

Fig. 8 - Modelling workbench a*ctive strut*

The active strut consists of a hydraulic cylinder with chamber (*hydraulicCylinder*), a hydraulic environment model (*hydraulicEnvironment*), and a servo-valve (*valve*) (see Fig. 8). Please note that the two undirected connections between the hydraulic cylinder and the servo-valve are no input/output relations; rather, this ideal hydraulic connection allows the user to pay no attention to the directions of pressures and volume flows.

To describe the individual aggregates, the user may pick one of two options: the experienced user can put in Objective-DSS directly via a text editor. A user who has hardly worked with the modelling workbench may have recourse to input assistants that facilitate access to the description language (see Fig. 9). The structure of the input assistants is the same with all parts and aggregates. In the upper part, there is the source code; the lower part displays the way the part or the aggregate is represented graphically. On the left-hand side, there are the buttons by means of which the source code is altered interactively.

Having set up the physical-topological model, one can automatically derive the mathematical model which will then be visualized by means of the concepts described above.



```
Servo43VoltageOdes named: ventilType.
  parameter: #(gain) on: ScalarOdes;
             #(d) on: ScalarOdes;
             #(tm) on: ScalarOdes;


    gain:                gain;
    gainVoltage:         1.0;
    damping:             d;
    timeConstant:        tm;

end.
```

Fig. 9 - Input assistant for servo-valves

7 OUTLOOK

To make the modelling workbench more efficient and comfortable to the user, there will be free forms for the graphical representation of topological-physical models. This means that there will no longer be block diagrams; instead, the wheel suspension, for instance, will be represented as a 2-D picture.

Moreover, an object-oriented database will take charge of the model management which currently is done in a file system. The management of models, parts, and aggregates will be extended in such a way that the different versions generated during development of a system will be managed as well. For this purpose a concept will have to be found that coordinates for instance versions of parts and aggregates.

REFERENCES

[1] Schröer, A.: Eine Modellbeschreibungssprache zur Unterstützung der Simulation und Optimierung von nichtlinearen und linearisierten hierarchischen Systemen, VDI Press, Series 20, No. 128, Düsseldorf 1994.

[2] Hahn, M.; Lückel, J.; Wittler, G.: Eine Entwurfsmethodik für Mechatronische Systeme. Magdeburger Maschinenbau-Tage: Entwicklungsmethoden und Entwicklungsprozesse im Maschinenbau, Magdeburg, 10. - 12. 9. 1997.

[3] Richert, J.; Hahn, M.: DSS - DSL - DSC: The Three Levels of a Model Description Language for Mechatronic Systems. Tampere Int. Conf. On Machine Automation, Tampere, Finland, 1994.

[4] Hahn, M.; Meier-Noe, U.: Objektorientierte Beschreibungsformen mechatronischer Systeme am Beispiel von DSS-Math. ASIM 94, 9. Symposium Simulationstechnik, Stuttgart, Oct. 10-13, 1994.

[5] Homburg, C.: Parallele Simulation mechatronischer Systeme auf der Basis von DSC, ASIM-Mitteilungen, Heft 50, Proceedings 10th Workshop „Simulation verteilter Systeme und paraleller Prozesse", Dresden, Germany, Oct.23-24, 1995.

[6] Hahn, M., and Wolf, M.: Klassenbibliothek: GRAPHEN, RINGE. Survey and documentation of the Smalltalk classes for Win32. MLaP, University of Paderborn 1995.

[7] Wolf, M.: Entwicklung und Implementierung eines graphischen Werkzeugs zu Analyse und Design modular-hierarchischer Systeme im Rahmen der objektorientierten Modellbildung mechatronischer Systeme (MOOMo). Diploma thesis, MLaP, University of Paderborn, 1995.

[8] Wolf, M., and Koert, D.: Klassenbibliothek: Placing & Routing. Survey and documentation of the Smalltalk classes. MLaP, University of Paderborn 1998.

[9] Booch, G.: Object-Oriented Design with Applications, Menlo Park, CA, 1992.

[10] Visual Smalltalk Enterprise, User's Guide, Parc Place: Digitalk Inc., USA, 1994.

[11] ADAMS simulation software for virtual prototyping of mechanical systems; Mechanical Dynamics, Ann Arbor, MI, USA, 1987

[12] Matlab, The Language of Technical Computing; Simulink, Dynamic System Simulation for Matlab; The Math Works, Natick, MA, USA, 1992.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

A development strategy for mechatronic systems based on functional and geometrical modelling techniques

M. A. Kümmel, A. Henke, J. Wallaschek

Heinz Nixdorf Institut, Universität-GH Paderborn, Fürstenallee 11, 33102 Paderborn, Germany

The complexity of modern mechatronic products is increasing permanently and is characterised by a distinct interaction of mechanics, electronics, control engineering and computer science. In order to minimise development time and costs while developing high-tech products, the use of systematic development methods is indispensable, but all existing methodic strategies are ineligible. This report suggests a new strategy for the development of mechatronic systems. It considers the development process from the product idea to the first functioning prototype and combines functional and geometrical modelling techniques. Besides the development of a typical mechatronic system is resumed.

1. INTRODUCTION

Modern high-tech products are becoming more and more complex regarding the use and combination of components from different engineering disciplines. Mechatronic products or systems (fig. 1) require a high integration level of mechanics, electronics, control engineering and computer science. Simultaneously the life cycles of the products are becoming shorter. Thus it becomes more difficult to develop high-tech products and still being in time and in cost.

One possible approach to improve the product development process is to use a methodic strategy. The strategy presented here bases on our experiences in the development of wire bonding machines, which are used in semiconductor manufacturing [1].

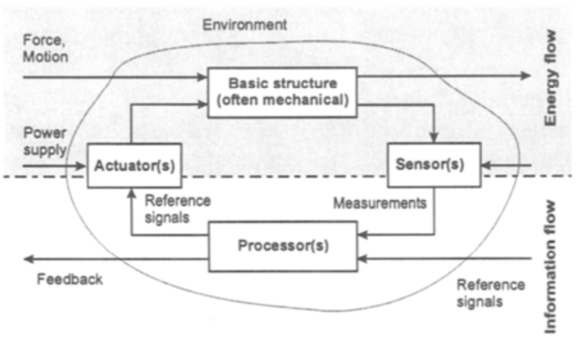


Figure 1. Basic structure of mechatronic systems.

2. EXISTING STRATEGIES

In the engineering disciplines, which are involved in the development of mechatronic systems, exist a lot of different methods and methodologies.

A general strategy for the design of mechanic systems acknowledged from most design methodology experts is described in the VDI-guideline 2221 [2], which is well known in Germany. This phase-oriented strategy is divided into seven working sections with seven corresponding results. These sections can be worked iteratively. Product specific, company specific or branch specific strategies can be derived from the general design strategy. Complete new product designs are well supported since the strategy emphasises early development stages. It also emphasises geometrical modelling techniques. Functional modelling techniques especially treating the dynamic system behaviour are regarded only slightly.

In electrical engineering miscellaneous design methods have been developed for special purposes. A strategy for the step-by-step design of analogue and digital circuits is described in [3]. This strategy is similar to the above mentioned VDI-guideline and consists of three phases with nine working sections. Here the analysis of the system behaviour and hardware tests, which strongly influence the following sections, are well considered. However this strategy is very product specific. In the area of digital circuits another strategy is often used, the hierarchical design [4]. This strategy differentiates between abstraction levels and domains. The development process is here defined as a series of transformations from one representation (specific level and domain) of a system to another. The direction in which the abstraction hierarchy is passed through defines the design type. The top-down design starts with the functionality of the system and ends at discrete components (geometry), the bottom-up design starts with standard components and ends at the functionality.

In control engineering the elaboration of development strategies is rudimentary, but many problem specific design and optimisation techniques have been created. A systematic approach for the design of control system is presented in [5]. There the processing of a control task is divided into seven sequential steps. This approach focuses on the systematic development of a control system. The design of the basic system is not included. Eventual improvements of the basic system which could result from the analysis of the control system are neglected.

In the field of computer science the investigation of strategy, phase and comparison models for software development is the main subject of software engineering. A very well developed phase model is the project model [6], which is subdivided into nine phases. These phases are logical units consisting of different activities and can overlap temporally. Iterations and feedback over several phases cannot be avoided in practise, but are neglected in this approach. In addition the user of the software is included in the design process only insufficiently. Improved cyclic strategies like the spiral model [7] have been developed with different approaches such as the so called prototyping or the evolutionary system design.

3. CHALLENGES

Five major challenges for a mechatronic development strategy result from the particular requirements of mechatronic systems and the inadequate existing development strategies.

3.1. Simultaneous engineering

Simultaneous engineering in mechatronics system development means corporate and parallel operation of all development engineers from different disciplines: mechanical engineers, electrical engineers, control engineers and software engineers. This approach requires a tight project management. A good teamwork is also very important to enhance communication and understanding between the diverse persons and to decrease the number of errors and iterations.

3.2. Integration of shape and function

The interaction of functional and geometric aspects in mechatronic systems is very high. It is not possible to look at function and shape separately. This becomes obvious especially at studying the dynamic system behaviour. Functional parameters like e. g. inertia, stiffness or electrical resistance depend intensely on the geometry and cannot be optimised only on the functional level.

3.3. Virtual prototyping

Virtual prototypes can be used to reduce expensive manufacturing of real prototypes. Computer internal models are synthesised with functional and geometrical modelling techniques. Static, kinematic and dynamic simulations as well as design studies provide early insight into the system. This information can also be useful to integrate other business units or customers.

3.4. Experimental validation

Even intense use of modelling and simulation could not reduce experimental investigations to zero. Virtual prototypes are only approximated and idealised representations of real systems. Experiments have to be done to identify system parameters and to prove the systems behaviour. Particularly at the development of new systems some properties such as damping or nonlinear material behaviour can only be estimated coarsely or even not be calculated.

3.5. Computer aided engineering

The selective use of proper modern software tools can improve the development process decisively. Various design and analysis tools like CAD software, FEM software, multibody-system software, circuit-analysis software, control system software or mathematics software are available. In addition the relevancy of tools which support conception and other engineering activities such as databases for principles and effects, project management software, documentation software, etc. is increasing.

4. DEVELOPMENT STRATEGY

The solution approach presented here tries to combine the advantages of existing strategies and to meet the above discussed challenges. This strategy bases on some assumptions which were listed as follows:

- The structure of the regarded mechatronic systems accords to fig. 1.
- The strategy covers a part of the complete product generating process, it starts with the product idea and ends at the first functioning prototype.
- The use of existing solution elements is predominantly considered for actuators, sensors and particularly processors.
- The regarded software covers mainly programs for closed and open loop control.

Another important assumption relates to the development organisation type. The development process should be performed as a project with a flexible project team. Typically this team consists of several engineers from mechanical, electrical, control and software engineering. All activities are performed in a corporate or parallel manner. During intense or difficult activities the core team can be extended with additional persons or be advised by experts.

The strategy plan for the development of mechatronic systems (fig. 2) consists of three phases: Conceptualise, Design and Analyse. These phases are not temporary separated units, but run fluently into one another. From a product idea respectively a real system to the first functioning prototype maximum eleven different working sections are passed through. It is possible to skip several sections, to go back to previous sections and if appropriate to process several sections in parallel.

In the conceptual phase the strategy distinguishes between a complete new development and a modification or improvement of an existing system which has not necessarily got to be a mechatronic system. Functional aspects play the main role at the first steps of a new product development and the creation working section has an essential meaning. The shape of the real system is the starting point of an improvement development, the conceptual phase consists only of the task clarification section. Nevertheless it can be determined in subsequent sections that a complete or partial new development of the system is necessary. In this case one has to jump back to the conceptual phase and to operate similar to a regular new development.

The activities and results of the particular working sections are described in detail in the following. Typical supporting software tools are listed briefly.

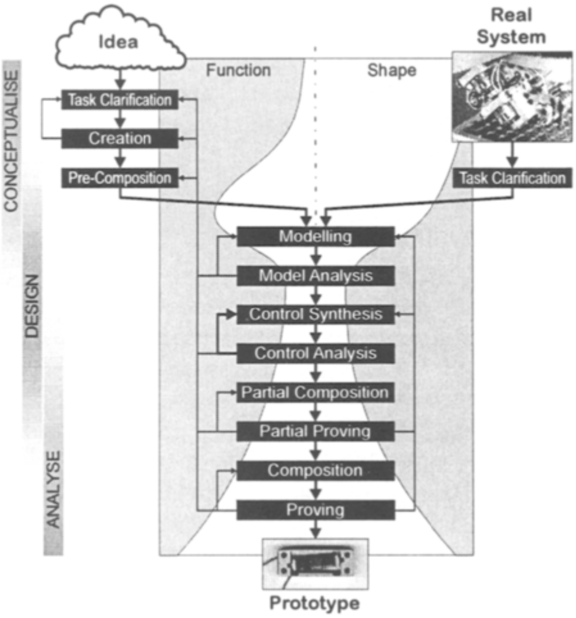


Figure 2. Mechatronic development strategy plan

4.1. Task clarification

In the first working section the product idea respectively the drawback of an existing system is precised in order to define a development goal accepted and apprehended by all participants. It has also to be clarified who does what, when, how, where and with which resources. The main working results are the system specification and the project management plan. The specification contains all necessary technical and product specific information for the realisation of the task and serves as evaluation scale. The project plan, which consists of several working activities and their personal, temporary and financial assignments, covers mainly organisational and economic aspects. Both working results have to be proved regularly and if necessary adjusted during the following development process.

Tools: documentation software, project management software.

4.2. Creation

Several sub-results were produced in this section. First of all the functional structure of the planned system has to be synthesised. It is build up with so called common functions, which are mostly described with a substantive and a verb. Especially in mechatronics this functional structure has to be extended with mathematical descriptions [8]. The mechatronic function notion additionally regards the kinematic, dynamic and control behaviour of the system. Often the functional structure cannot be synthesised with all these additions a priori, but it should grow during the development project.

In the next step principles and effects fulfilling the partial functions are searched. There exist miscellaneous creative methods supporting this search process such as brainstorming, method 6-3-5 or lateral thinking. In addition construction catalogues, patent recherches or product catalogues can be used. On the basis of suitable principles and effects alternative solutions are then created. Thereby certain principles or effects can be substituted with existing solution elements (screws, transistors, electromagnets, microprocessors, etc.).

The different solutions were evaluated subsequently. Simple point scoring methods, efficiency analysis or similar methods can be used for evaluation and selection of the most suitable solution variant. The selected solution which is the main result of this working section is denoted as the concept. Depending on the current task it can be reasonable to pursue several concepts.

Tools: sketch software, principles and effects database software, evaluation software.

4.3. Pre-composition

The main activities of the pre-composition section are the detailing and preparation of the concept for the following modelling section. This includes the coarse dimensioning of several components as well as the adjustment to selected solution elements. Working result is a pre-draft of the system from which all important system parameters can be determined or estimated.

Tools: mechanical and electrical CAD software.

4.4. Modelling

Modelling is a multistage working section [9] which is intensely connected to the next three sections. It leads to the so called model level where computer aided simulations of the system behaviour can be performed.

First of all a mental model is derived from the pre-draft respectively the real system with respect to the development task. An idealised image of the system is developed by using suitable assumptions and simplifications as well as neglecting and omitting of nonrelevant details. In the next step the mathematical equations necessary for specifying the processes inside the system are derived. On the level of the mathematical model it is possible to describe mechatronic systems uniformly and independently from engineering discipline descriptions. Subsequently the equations were adjusted to numerical solving methods. The result of this adjustment is denoted as the algorithmic model. Finally this algorithm is converted to an executable computer program with the use of a programming language.

The main task of the participating engineers in this section is the deriving of the mental model. All other models particularly both last mentioned can automatically be synthesised today.

Tools: mathematics software, multibody-system software, circuit-analysis software, FEM software.

4.5. Model analysis

The system behaviour is simulated and analysed using the previously generated computer model in this working section. For example static, kinematic, dynamic, thermal, electric or magnetic analyses of the system are processed. The simulation results do not only influence the following sections, but are also used for model validation or can require an iteration of previous working sections.

Tools: mathematics software, multibody-system software, circuit-analysis software, FEM software.

4.6. Control synthesis

In this working section the control system concept that bases upon the results of the model analysis is developed and the dynamic control behaviour is optimised regarding the system specification. For this purpose a lot of diverse mainly problem specific synthesis (frequency response method, compensation control, pole selection, etc.) and optimisation methods (quadratic control area, absolute value optimum, etc.) are available. This section and the following section are often processed multiple.

Tools: control system software, mathematics software.

4.7. Control analysis

The main activity of this working section is the investigation of the control system behaviour. The system answer is evaluated with suitable criteria such as rapidity, accuracy or disturbance sensibility. The results of this investigation are compared with the system specification. If a sufficient system behaviour cannot be obtained, the conceptual phase has to be passed through again to alter the system parameters properly.

Tools: control system software, mathematics software.

4.8. Partial composition / Composition

Supplementary details and missing elements are added to the complete system design respectively to parts of the system design in this working section. Also the connections between the several components have to be determined. Manufacturing instructions for a prototype system are the working results. It is possible to realise only partial components of the prototype. Then all other essential components have to be simulated with a computer model.

Tools: mechanical and electrical CAD software, programming software.

4.9. Partial proving / Proving

One of the essential aspects of the development of mechatronic systems is the proving respectively partial proving of the system. This covers not only investigations for the verification of the system behaviour, but also validations of the system models as well as the experimental determination of system parameters. Hence this section can influence the control synthesis and modelling sections as well as sections of the conceptual phase.

Tools: test and measurement software, hardware in the loop software, experimental modal analysis software.

The final result of the mechatronic development process is a functioning prototype, which can be developed subsequently to a competitive product with all specifications for manufacturing, assembly, inspection, transport and use.

5. EXEMPLARY PRODUCT DEVELOPMENT

Aluminium wire bonding machines are used in semiconductor industry to generate the electrical connections between chip and board. One important task of these machines is the accurate and fast positioning of the aluminium wire. In present bonding machines the wire feeding system is driven by electromagnetic actuators where the positions are defined by mechanical stops.

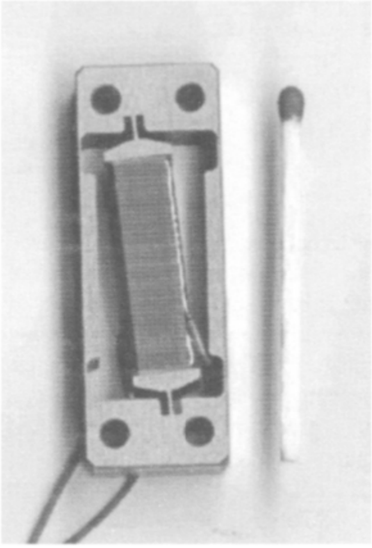


Figure 3. Piezoelectric actuator for a wire feeding system.

The main goals of the development of a new wire feeding system are programmable positions, higher positioning speed and accuracy as well as compact design. Piezoelectric actuators are predestined for such purposes. Thus several solutions with piezoelectric actuators have been investigated in the conceptual phase. Computer aided analyses with multibody system and FEM software of coarsely designed solutions have been used to support the selection of the most promising concept. In addition first control concepts have been investigated. Finally the combination of a multilayer actuator with a parallelogram displacement amplifier has been chosen for subsequent elaboration.

Further FE analyses of the system have been processed to optimise the design of flexure hinges, to calculate transient responses and dynamic modes. A simplified multibody model of the system has been generated with the analysis results. This dynamic model was used to synthesise an open loop control strategy and to optimise the design of the electronic power amplifier. In parallel a prototype of the system was designed and manufactured (fig. 3).

Experimental investigations of the prototype have been processed to validate system parameters and to optimise the control concept. The system has a maximum stroke of 300 μm with a positioning time of less than 4 ms and an accuracy of less than 5 percent. The outer dimensions of the system are 48×17×8 mm^3. This system is about five times faster and smaller than the present system. In addition the positions can be comfortably programmed.

6. CONCLUSION

In this paper a systematic approach was proposed for improving the development process of mechatronic systems. It bases on the combination of existing functional and geometrical modelling techniques and tempts to meet five major challenges: simultaneous engineering, integration of shape and function, virtual prototyping, experimental validation and computer aided engineering.

In the future the efficiency of the presented strategy has to be validated with further suitable product development processes. Furthermore the structuring of large systems into small subsystems has to be considered.

REFERENCES

1. Henke, A.; Kümmel, M. A.; Wallaschek, J.: A piezoelectrically driven wire feeding system for high performance wedge-wedge-bonding machines, Proceedings of Mechatronics '98, 1998.
2. VDI-Richtlinie 2221: Methodik zum Entwickeln und Konstruieren technischer Systeme und Produkte, Beuth Verlag, 1993.
3. VDI/VDE-Richtlinie 2422: Entwicklungsmethodik für Geräte mit Steuerung durch Mikroelektronik, Beuth Verlag, 1994.
4. Armstrong, J. R.; Gray, F. G.: Structured logic design with VHDL, Prentice-Hall, 1993.
5. Föllinger, O.: Regelungstechnik, Hüthig, 1994.
6. Hesse, W.: Handbuch der Informatik, Band 5.3: Software-Entwicklung, Oldenbourg, 1992.
7. Boehm, B. W.: A spiral model of software development an enhancement, Computer, May 1988, pp. 61-72.
8. Lückel, J.; Wallaschek, J.: Functional modeling and simulation in mechanical design and mechatronics, 2nd MATHMOD/IMACS Symposium Mathematical Modeling, 1997.
9. Wallaschek, J.: Modellierung und Simulation als Beitrag zur Verkürzung der Entwicklungszeiten mechatronischer Produkte, VDI-Berichte 1215, VDI-Verlag, 1995.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

3-D Graphical Simulation for Agile Modular Machine Systems

Josef Adolfsson*, Petter Olofsgård*, Philip Moore+, Junsheng Pu+

*Centre for Intelligent Automation, Department of Engineering Science University of Skövde, PO Box 408, 541 28 Skövde, Sweden

+Mechatronics Research Group, Faculty of Computing Science and Engineering De Montfort University, Leicester, LE1 9BH, UK

This paper presents a system architecture for a 3D graphical simulation system for the design, simulation and off-line programming of modular machine systems adopting a distributed control architecture. The simulation system aims to support the application engineer to shorten the design to implementation process of modular machine systems. The simulation system aims to support development of individual control of each machine elements, sensor interaction, and the communication between the modules within the machine system. The system will also enable real-time interaction with the real machinery for evaluation of minor changes in the machine system, hence the link between the design/simulation environment and the run-time environment is an important aspect of the work.

1. Introduction

Global competition in manufacturing and changing customer demand are resulting in a trend towards greater product variety and innovation, shorter product life-cycles, lower unit costs and higher product quality[1]. To cope with these demands companies must facilitate production systems that accommodate higher flexibility, reduced lead-times for the design to implementation process and increase the re-configurability of the manufacturing capability. The common technological approach adopted to achieve flexibility has been to use systems based upon re-programmable equipment e.g. industrial robots, CNC-machines, etc.. [3]. One approach to achieve enhanced flexibility of production facilities is to build machine systems from standard components e.g. mechanical components, sensor components, control components, etc.. [4][1]. Such a machine system can be modified or upgraded as required, to suit the task on hand, and thereby resist obsolescence, shorten redesign and enabling new designs to be realised by utilising existing modules. In order to achieve total system modularity, also the control system has to be modularised [2]. Distributed control e.g. fieldbus technology is a way forward to modularise control systems. The use of a distributed control system has a number of advantages compared to the centralised control system such as:

i) reduced software complexity

ii) reduced coupling among control components

iii) Implicit fault tolerance

iv) Intelligence can be distributed to the component/ device level of the system and thereby support the component based methodology[11][5][6].

Each component can then be seen as an autonomous entity within the machine system which communicates with other components, for synchronisation purposes, via a control network. The difference between a centralised and a distributed control system is shown figure 1 and figure 2.

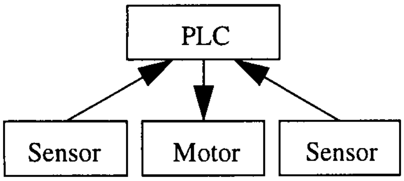


Figure 1. Centralised control system

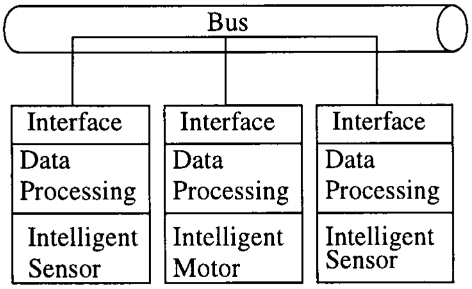


Figure 2. Distributed control system

2. Virtual Environment

The demand for reduced lead-times results from increased competition forcing manufacturing companies to realise new products more rapidly. In combination with shortening life-cycles of these products, the lead-times for the development and changes to related manufacturing have to be continually reduced. Furthermore, in many instances it is no longer economically viable to produce completely new manufacturing facilities for every new product introduced [4]. One way to reduce the lead-time for the design process is to utilise graphical modelling and simulation tools in 3-D CAD environments in the engineering and development process.

Graphical modelling and simulation in three dimensional CAD environment, an example is shown in figure 3, can provide the basis for 'virtual manufacturing', an example of which is 'Computer Aided Robotics' [7][8]. CAR systems provide facilities for evaluating different automation design and equipment capabilities. For example, they support investigations to optimise the choice of robot and auxiliary tooling for a planned task, prior to any commitment of capital [9]. Furthermore, the same environment can be used for program evaluation and code generation thus shortening the design to manufacture cycle [10]. CAR systems can also be used in the development phase for special purpose automation systems such as modular handling systems [Craig].

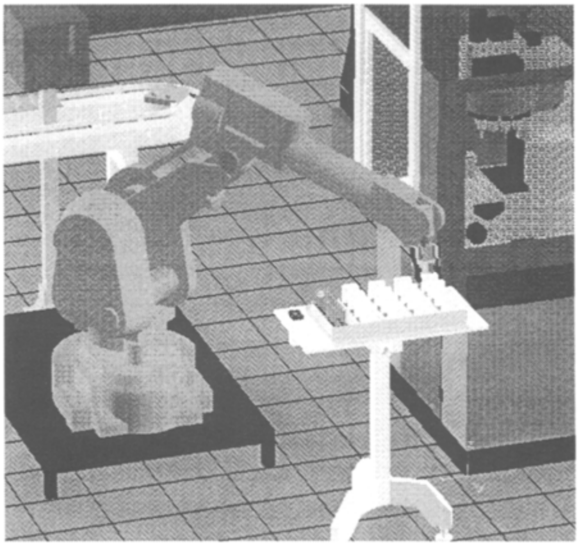


Figure 3. Simulated robot workcell

'Virtual environments' provide an ideal platform to facilitate modular machine design where each system is likely to be a one-off 'customised design' utilising proven components. However, a systematic way of creating and developing 'virtual components' including both control and sensing aspects is lacking. Moreover, methods for the seamless linking of the 'virtual environment' and the 'real environment' have not been reported. The importance of these capabilities can not be over emphasised in realising the virtual engineering of modular machine systems.

3. Modular Machine Design Environment (MMDE)

The MMDE will be implemented in Robot Studio, a robot simulation system from Prosolvia Systems. The MMDE will provide the user with functionalities like design, simulation, analysis and off-line programming of machine systems which are controlled by a distributed control architecture and is based upon a component based paradigm. The system should as much as possible reflect the structure and the behaviour of the real machinery, therefore, a number of functionalities for the system has been identified:

i) design of mechanical components, such as linear and rotary actuators, with associated behaviours. Hence both the CAD model of the component and its associated control logic and parameters can be developed within the same environment. The control code developed in the virtual environment can then be used to drive the real component.

ii) design and simulation of sensor components. Decisions taken in a machine system are in many cases determined by events triggering sensor signals, therefore sensor interaction must be represented in the simulation environment.

iii) design of a machine system by logically aggregating groups of actuator components, sensor components, etc..

iv) Simulation of the communication between the components.

v) logical description and analysis of the complete machine system / production facility. One powerful way to analyse the control of a system is to observe its behaviour [15]. A 3-D virtual environment can provide a suitable tool to visualise the functionality of the machine system for the engineers, and

vi) off-line programming of the system. All control code developed in the simulation environment can after debugging and optimisation be transferred to the real machinery.

3.1. Design levels

The system architecture for the modular machine design environment is shown in figure 4. Four different levels of design issues has been identified to fullfill the requirement of the system

3.1.1. Design of a new component

The design of new components will be supported by the use of templates, that will guide the user through the design process. There are two aspects of defining a component, modelling of the components in terms of geometrical and kinematic properties and assign attributes to the components i.e. attributes that can be obtained from the suppliers specification of the product. A library of created components will be accessible through the component bus.

3.1.2. Design of components with behaviour

A component with behaviour is basically obtained by combining an actuator or sensor component with a control component. The behaviour and how the components interact, with the rest of the system, is defined by the control logic in the control component. The code developed in the simulation environment would be used to drive both the simulation and the real component, after suitable postprocessing.

3.1.3. Machine system design

Designing a machine system incorporates both the design of the physical structure and the design of the logical structure of the system. The physical structure is defined by combining previously designed components and components with behaviour into a machine system. Whereas the logical structure is defined by:

i) specify the connections between the control components i.e. how the control components communicates over the network:

ii) define the supervisory control logic for the whole system i.e define the behaviour of the whole system

3.1.4. System simulation and analysis

The connections between components that are defined in level 2 is in this stage transformed to a "virtual fieldbus" which simulates the network in the real machine. All control components in the virtual environment can read and update the information in the virtual fieldbus. The status of the virtual fieldbus is used to drive the simulation, since it contains information about sensor signals, status of actuators etc.. A simulation in 3D is very powerful tool for the designer to evaluate the behaviour of the system. Therefor, errors in the control logic can in an early stage be detected and corrected.

4. Information Infrastructure

To be able to design, simulate, and analyse the modular machine systems and it's distributed control, the 'virtual environment' need to incorporate a coherent IT supporting environment, based upon the object-oriented paradigm[16]. An information infrastructure and associated integration services has to be established to handle:

i) the actual components or templates of designed components for later retrieval when needed by the users and for component re-use

ii) the actual code that is produced by the Virtual Environment, which also needs to be accessible by the run-time environment

iii) the information needed exclusively in the Virtual Environment to run and debug the simulation. Components state at a certain moment, but also previous states, debug information like breakpoints, etc.

iv) the information that is stored together with a component that will help the user of the Virtual Environment to insert that component when building a modular machine system

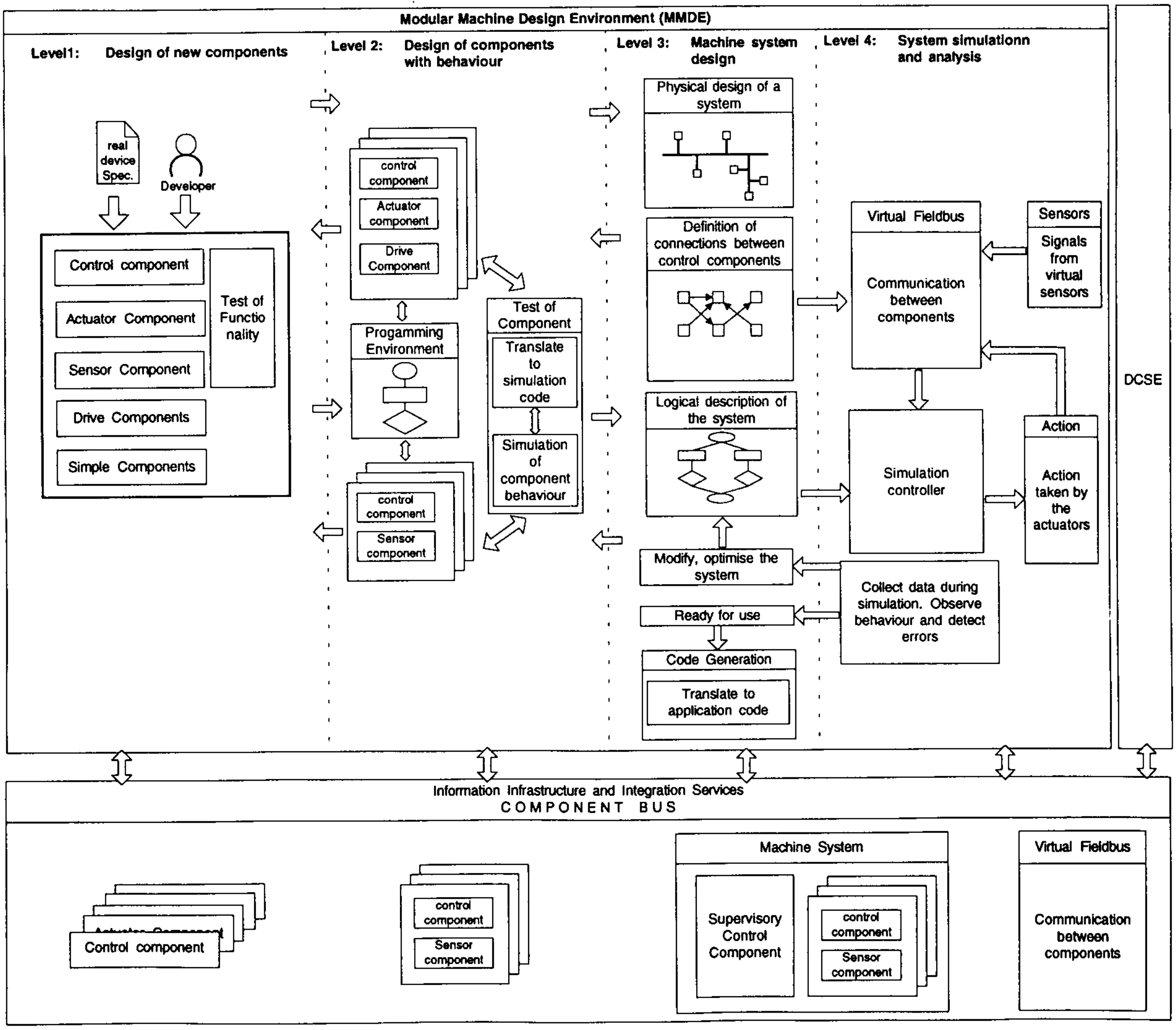

Modular Machine Design Environment (MMDE)
Level1: Design of new components
Level 2: Design of components with behaviour
Level 3: Machine system design
Level 4: System simulationn and analysis
real device Spec.
Developer
Control component
Actuator Component
Sensor Component
Drive Components
Simple Components
Test of Functionality
control component
Actuator component
Drive Component
Progamming Environment
control component
Sensor component
Test of Component
Translate to simulation code
Simulation of component behaviour
Physical design of a system
Definition of connections between control components
Logical description of the system
Modify, optimise the system
Ready for use
Code Generation
Translate to application code
Virtual Fieldbus
Communication between components
Sensors
Signals from virtual sensors
Simulation controller
Action
Action taken by the actuators
Collect data during simulation. Observe behaviour and detect errors
DCSE
Information Infrastructure and Integration Services
COMPONENT BUS
Actuator Component
Control component
control component
Sensor component
Machine System
Supervisory Control Component
control component
Sensor component
Virtual Fieldbus
Communication between components

Figure 4 System architecture

Other functions that are needed to handle the information are version control, transactions, and access restrictions.

The information infrastructure and associated integration services is realized through a "component bus". The idea of the component bus is to achieve a higher level of abstraction, se figure 3, than regular low-level access to databases or similar. The component bus contains all information associated with a component and to the virtual environment, it looks like there is an actual component within the component bus, the virtual environment then can make a request to the component bus to be able to access the component. All access to the component is done through well defined interfaces. Since the object-orientation paradigm will be used to structure the components, the components bus will be implemented through the use of Distributed Object Computing

This information infrastructure and integration service will be the information kernel that will be accessible by several users, running the Virtual Environment, and the run-time environment.

5. Link to Real Machine

All information exchange between the virtual environment and the run-time environment will be done through the component bus. This means that there must exists application on both sides which should be able to handle the exchange of necessary information concerning the interaction between the virtual environment and the run-time environment

6. Future Work

The proposed system architecture will be implemented in the robot simulation software PS-Engine. A small modular machine system will be used for evaluation of the design and simulation system. The machine system is based on a distributed control architecture and consists of 3 standard linear actuators, that are combined into a 3DOF Cartesian robot, a conveyor system and proximity sensors. The work is a part of the ESPRIT funded project: Integrated Design, Simulation and Distributed Control of Agile Modular Manufacturing Systems" and will therefor be implemented in the project's demonstrator cell, a system for flexible assembly of automotive aggregate units.

Conclusion

The use of a virtual environment, for building modular machine systems, can dramatically reduce machine / system development lead-times since a lot of the proof-of-concept, selection, evaluation and development can be done before the real machine is ordered, assembled and commissioned. In additional, different configurations of the machine system can be tested to find the optimal solution, without access to the real plant or equipment.

Acknowledgments

The work described in this paper is a part of the research project "Integrated Design, Simulation and Distributed Control of Agile Modular Machine Systems" and is funded under the Framework IV ESPRIT programme and constitutes part of the Integration in Manufacturing initiative. Project partners include; Baldor Optimised Control; De Montfort University; Echelon; Honeywell Control Components; J A Krause; Prosolvia Systems; University of Skövde and Volvo Automation

References

1. Pu J. and Moore P R "Component/Image-based Design of Distributed Manufacturing Machine Control Systems", ICRAM'95, Istanbul, Turkey, August, 1995
2. J. Wikander, M. Törngren "Decentralized control systems for modular machines" Proc. of the International Conf. on Industrial Electronics, Control and Instrumentation, Bologna, 1994, pp. 1490-1994
3. R. Cohen, M. G. Lipton, M. Q. Dai and B. Benhabib, "Conceptal Design of a Modular Robot", Journal of Mechanical Design, March 1992, Vol. 114/117
4. G. G. Rogers and L. Bottaci, Modular Production Systems: A New Manufacturing Paradigm, Journal of Intelligent Manufacturing (1997), 8, pp. 147-156

5. D.M. Dilts, N.P. Boyd and H.H. Whorms, The Evolution of Control Architectures for Automated Manufacturing Systems, Journal of Manufacturing Systems, Volume 10/No. 1, pp. 79-93
6. Neil A. Duffie, Ramesh Chitturi, Jong-I Mou, Fault-tolerant Heterarchical Control of Heterogeneous Manufacturing System Entities, Journal of Manufacturing Systems, Vol. 7, No.4, pp. 315-327.
7. Benhabib B, Zak G. and Lipton M G., "A generalised kinematics modelling method for modular robots" J. of Robotics Systems, 1989, 6(5), pp 545-571
8. Craig J, "Issues in the design of Off-line programming Systems Robotics Research", MIT Press, 1988, ISBN 0-262-02272-9
9. Bien C., "Simulation a necessity in safety engineering" Robotics World December, 1992, pp 22-26
10. Renfors J, Aalto H, and Jokela M, "Accurate off-line programming for a 12-DOF Robot Cell at Bronto Skylift Ltd", Proceedings Robotikdagar 2-3 June 1993, Linköping, ISBN 91-7871-136-3
11. C. Xie, J. Pu and P.R. Moore "A case study on the development of intelligent actuator components for distributed control systems using LONWORK Neuron Chips", International Journal of Mechatronics
12. Ravani B., "World Modelling for CAD-based Robot Programming and Simulation", CAD Based Programming of Sensory Robots; Springer Verlag, 1988
13. Smith M G, "An environment for more easily programming a robot", Proceedings of the IEEE International Conference on Robotics and Automation, Nice, France, May 1992, pp 10-16
14. Eriksson P T and Moore P R, "Improved Computer Aided Robotics with simulation of sensors" Proceedings Mechatronics and Machine Vision in Practise 1994, IEEE Computer Society Press, ISBN 0-8186-6300-6
15. Ju-Yeon Jo, Yoohwan Kim, Andy Podgurski, and Wyatt S. Newman "Virtual Testing of Agile Manufacturing Software Using 3D Graphical Simulation" Proc. of the conf. IEEE ICRA'97, 1997
16. H.Tsukune, M. Tsukamoto, T. Matsushita, F. Tomita, K. Okada, T. Ogasawara, K. Takase, and T. Yuba "Modular manufacturing" Journal of Intelligent Manufacturing (1993), pp 163-181

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Object-Oriented Modeling and Simulation of Mechatronic Systems with 20-sim 3.0

P.B.T. Weustink, T.J.A. de Vries, and P.C. Breedveld

Drebbel Institute for Systems Engineering
Department of Electrical Engineering, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
Phone: +31-53 489 28 06, Fax: +31-53 489 22 23
E-mail: mechatronics@rt.el.utwente.nl, www-address: http://www.rt.el.utwente.nl/mechatronics

Keywords: mechatronics, object-oriented modeling, conceptual design, modeling and simulation

1 INTRODUCTION

There is an increasing need for controlled electro-mechanical systems with more flexibility, higher performance and reliability. It has been recognised that these systems are the result of design activities of not just one discipline, but of several disciplines [8]. This lead to the introduction of *mechatronics*, a design approach that combines mechanics with electronics and information technology to form both *functional interaction* and *spatial integration* in components, modules, products and systems.

This approach differs from classical design patterns for controlled electro-mechanical systems, in which the design starts with the mechanical subsystems, next deals with the electrical subsystems and finally the controllers, that are designed to obtain the specified performance [1].

1.1 Mechatronic design

De Vries [2] described specific problems that occur during mechatronic design and stated that "a major cause of the problems is the lack of proper support for the *conceptual design task*". In the conceptual design stage a rough idea is developed of how the project will function and what it will look like. As the conceptual design stage is an early stage in the design process it is well suited to determine the functional interaction between the different subsystems.

In general, the design of a mechatronic system will be a complex problem. Separating partial problems is a possibility to tackle such a complex design problem. In Simon's words [3]: "the design problem is ill-structured in the large, but well structured in the small". Hence the complex ill-structured design problem should be split into small well-structured problems to which a local problem solver can be applied.

These well-structured problems should not just be considered in a specific domain, as this limits the solution space too strongly. Domain specific sub-problems should be solved by taking into account the consequences of a solution in other domains or by finding alternative solutions in other domains, as seen in figure 1.

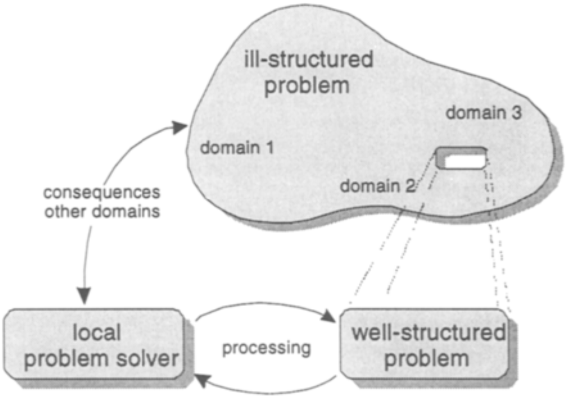


Figure 1: Schematic structure of solving ill-structured problems

1.2 Object-oriented modeling

Modeling, simulation and analysis tools are needed to support the mechatronic designer during these activities [2]. In the conceptual design stage these tools should consider the functional interaction between domain specific subsystems. There are many tools available that support modeling of *dynamic systems*, like 20-sim [5], Simulink [6], and Dymola [7], but the models all reside in the information domain (and the energy domain in 20-sim). This paper shows that *mechatronic systems* require an object-oriented modeling approach using the energetic domain information. The environments (i.e. 20-sim) should be changed in several ways to improve the support for this.

2 MECHATRONIC SYSTEM EXAMPLE

As an example of the approach a simple laboratory setup called *Linix* is used, shown in figure 2.

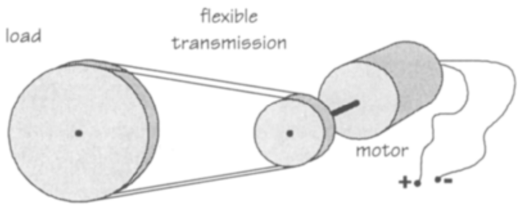


Figure 2: Sketch of the Linix laboratory setup

It consists of a computer-controlled actuator that drives a load using a flexible transmission. The power amplifier and the digital controller part of the setup are not shown here. The system is representative for a large class of mechatronic systems, since its dominant behaviour is that of a 4th-order system [1].

2.1 Model of the system

The top-level model is shown in figure 3. It is intrinsically based on *power exchange* between the connected submodels. For instance, the *motor* is a submodel with electrical power input and mechanical power output. This submodel can be decomposed into idealised elements as shown in another iconic diagram (figure 4). In this diagram the transducer is an elementary submodel described with simple gyrator equations:

$$\begin{aligned} T &= k_m \cdot I \\ V &= k_m \cdot \omega \end{aligned} \tag{1}$$

2.2 Object-oriented model characteristics

The model described above complies with a general model structure given in figure 5. Models of dynamic systems will generally have such a structure. The general model has the following characteristics:

- States, parameters, inputs and outputs are incorporated in submodels (*encapsulation*).
- The internal behaviour of a submodel is described as a composition of lower level submodels or a set of equations (*part-of model hierarchy*).
- A structure of interconnected sub-models can be specified without explicitly determining the computational causality (*non-causal modeling*).
- The instantiation of a submodel is separated from its definition in a library (*abstraction*).
- In the submodel definition, a separation can be made between externally accessible attributes (ports, parameters, states) and internal behaviour.

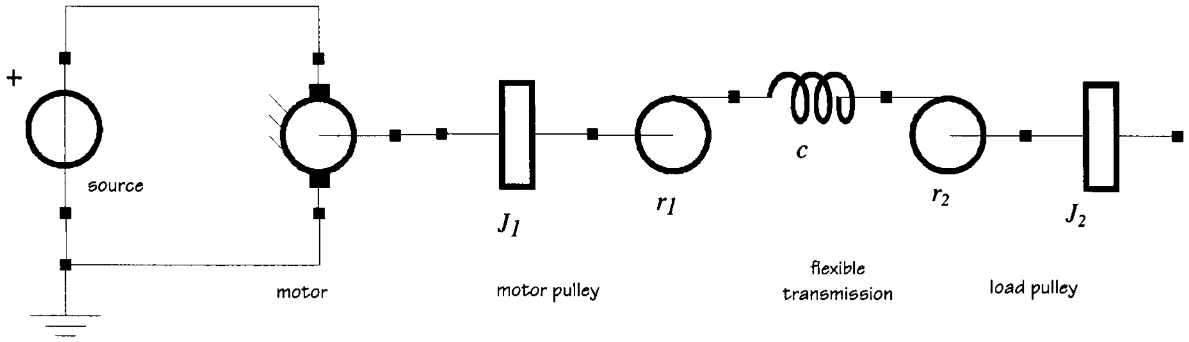


Figure 3: Iconic diagram of the system

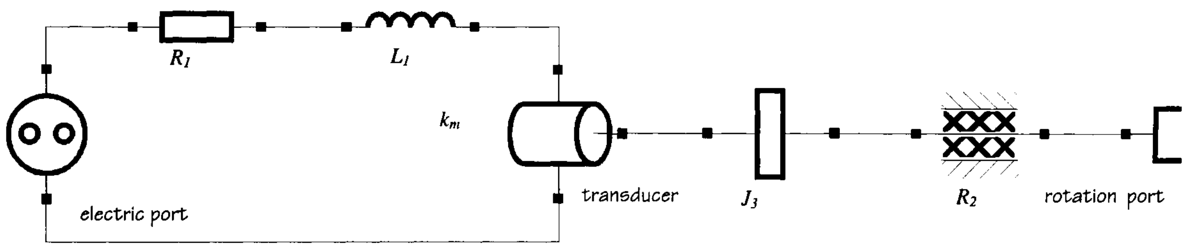


Figure 4: Composite submodel of the motor

Multiple versions for the internal behaviour may exist (*polymorphism*).

- Submodel definitions can be described as an extension of a more general submodel definition (*kind-of hierarchy in the submodel library*).

In our view, these are the characteristics that an *object-oriented model* should have. When using an object-oriented approach, modeling now becomes the selection or creation of relevant submodels and the determination of their interconnections [2]. The structure of an object-oriented model is given in figure 5.

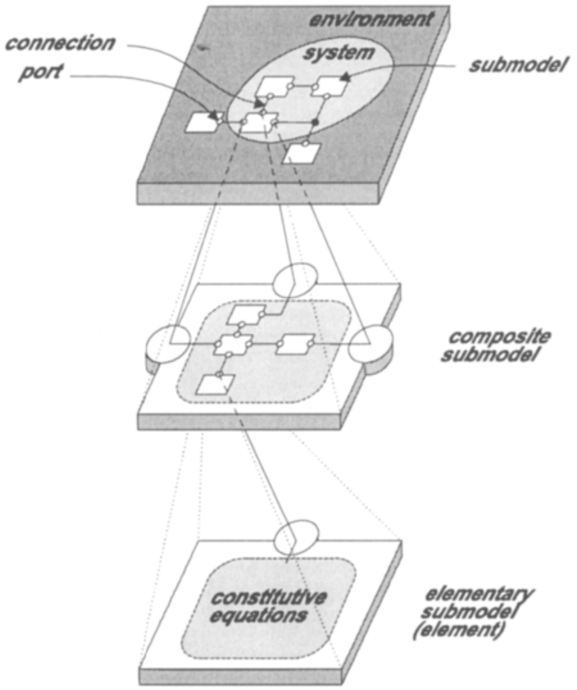


Figure 5: Structure of an object-oriented model

3 COMPUTER-BASED SUPPORT

A modeling and simulation tool for mechatronic systems should be able to handle object-oriented models in multiple physical domains, since the model of a mechatronic system consists of the integration of submodels in multiple physical domains and the information domain.

Tools should have a modeling technique that explicitly describes the *energetic behaviour* of the mechatronic system, and should have a proper support of the modeling process. Modeling itself is a dynamic process during which the kind of submodel often changes. In the Linix example the modeler could start with just a motor submodel and change it later into a brushless DC motor, for instance, when more is known about the design.

It is the problem context that determines the implementation of the submodel. The same kind of submodel has several different implementations. In the example, the motor submodel is implemented as an iconic diagram of elementary submodels, whereas in another problem context a complex set of equations is required. Proper support would enable the exchange of submodel implementations with minimal effort.

4 LINIX EXAMPLE IN 20-SIM 3.0

The 20-sim environment has been completely rewritten into a new version to implement the mentioned features. The bond-graph input of the Linix example is shown in the editor in figure 6. At the left-hand side, the complete *part-of* model hierarchy is shown (as described in figure 5). The implementation of the selected submodel in this hierarchy determines the kind of editor that is used

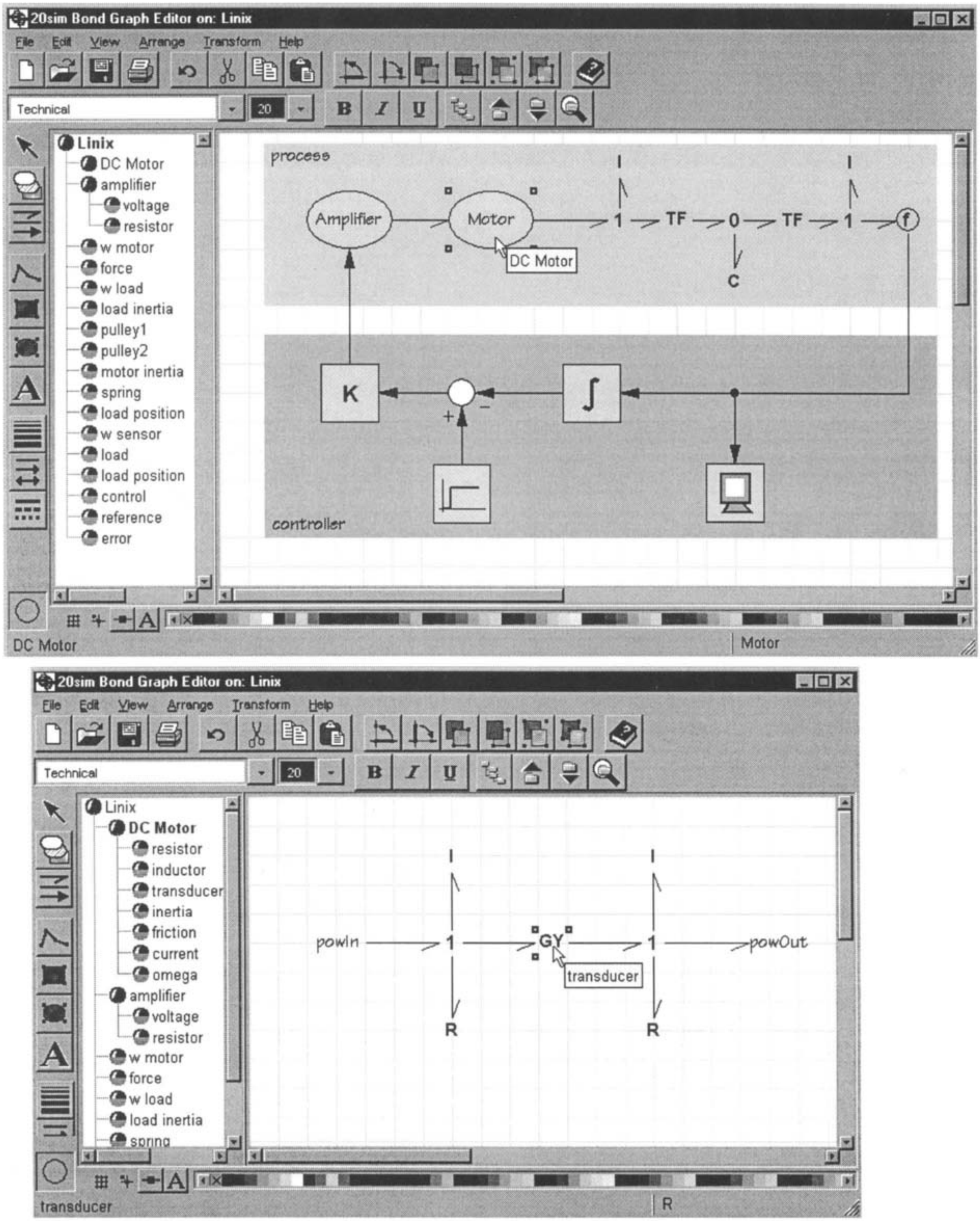
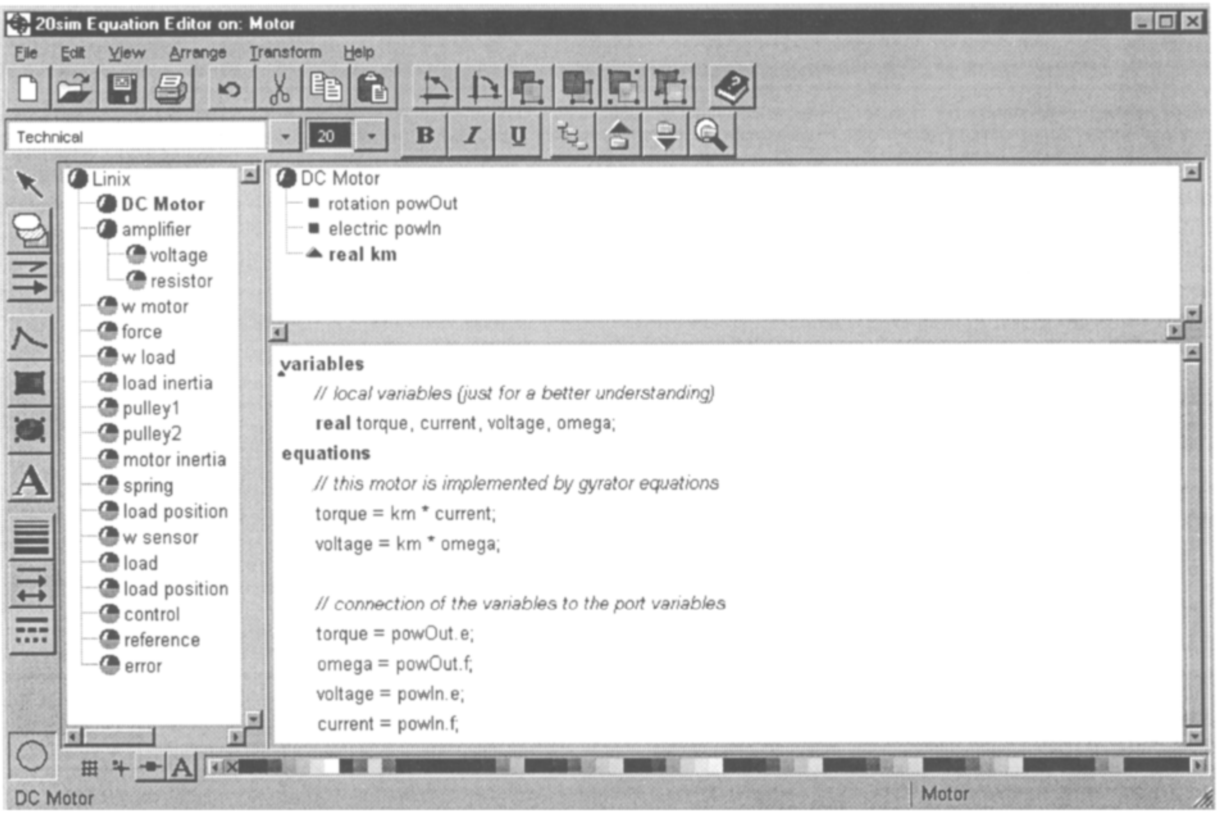

20sim Bond Graph Editor on: Linix
File
Edit
View
Arrange
Transform
Help
Technical
20
Linix
DC Motor
amplifier
voltage
resistor
w motor
force
w load
load inertia
pulley1
pulley2
motor inertia
spring
load position
w sensor
load
load position
control
reference
error
process
Amplifier
Motor
DC Motor
I
1
TF
0
TF
1
f
C
K
+
−
controller
DC Motor
Motor
20sim Bond Graph Editor on: Linix
Linix
DC Motor
resistor
inductor
transducer
inertia
friction
current
omega
amplifier
voltage
resistor
w motor
force
w load
load inertia
spring
powIn
1
GY
1
powOut
I
I
R
R
transducer
transducer
R

Figure 6 and 7: The 20-sim bond graph editor with the top-level model and the motor submodel in the right hand part of the window. In figure 6 the top-level model *Linix* is selected. The implementation is a bond graph of the process, together with a block diagram that represents a controller of the angular position of the load. Both the *DC motor* and the *amplifier* submodels have an implementation, the other submodels only have a type description in order to demonstrate that the tool supports a top-down approach. The *motor* submodel is implemented as a bond graph containing the parts shown in figure 4 as ideal bond graph elements. Figure 7 shows this implementation.

Next the *motor* submodel changed with a view mouse clicks from a bond graph implementation to a set of equations. The equation editor is now used to represent the implementation of the selected submodel (figure 8). The upper part of the equation editor shows the type information, in this case the two ports and the type parameter *km*. The lower part is for entering and modifying the equations in the SIDOPS+ language [4]. The equation editor gives the user feedback through extensive equation analysis and feedforward through colour syntax highlighting.

When the *amplifier* submodel is selected in the hierarchy, the editor subwindow changes into an iconic diagram. The implementation of the amplifier is shown in figure 9. The amplifier has an input signal *m* and an electric port *p*, represented by two nodes, *p high* and *p low*. Ports and signals are represented in the diagram by generic plugs.

When the model is implemented completely, it is translated into a set of simulation instructions that is performed by the simulator. Since it is one environment, there is shared memory between the simulation kernel, plot package and the model editors. This enhances model consistency, and enables submodel debugging and editor updates during simulation.


20sim Equation Editor on: Motor
File
Edit
View
Arrange
Transform
Help
Technical
20
B
I
U
Linix
DC Motor
amplifier
voltage
resistor
w motor
force
w load
load inertia
pulley1
pulley2
motor inertia
spring
load position
w sensor
load
load position
control
reference
error
DC Motor
rotation powOut
electric powIn
real km
DC Motor
Motor

```
variables
    // local variables (just for a better understanding)
    real torque, current, voltage, omega;
equations
    // this motor is implemented by gyrator equations
    torque = km * current;
    voltage = km * omega;

    // connection of the variables to the port variables
    torque = powOut.e;
    omega = powOut.f;
    voltage = powIn.e;
    current = powIn.f;
```

Figure 8: The 20-sim equation editor

5 CONCLUSIONS

The new version of 20-sim features object-oriented models, thus including:

- multi physical domain and the control domain
- polymorphic models
- multiple model representations
- possible transformations between these
- libraries with submodels organized in a *kind-of* hierarchy (with inheritance of relevant properties)

These features make 20-sim 3.0 suited for mechatronic system design. Future work resides on the development of more domain specific libraries with iconic diagram representations.

REFERENCES

[1] Coelingh, H.J., T.J.A. de Vries and J. van Amerongen (1998), *Automated Performance Assessment of Mechatronic Motion Systems During the Conceptual Design Stage*, to be presented at ICAM 1998, Okayama, Japan.

[2] Vries, T.J.A. de (1994), *Conceptual design of controlled electro-mechanical systems*, Ph.D. thesis, University of Twente, Netherlands

[3] Simon, H.A. (1973), *The structure of ill-structured problems*, Artificial Intelligence 4, 181-20.

[4] Breunese, A.P.J. (1997), *Automated Support in Mechatronic Systems Modeling*, Ph.D. thesis, University of Twente, Enschede, Netherlands.

[5] Controllab Products (1997), *20-sim Reference Manual*, Enschede, The Netherlands, http://www.rt.el.utwente.nl/20sim

[6] The Mathworks Inc. (1998), *Simulink 2.2*, http://www.mathworks.com/products/simulink

[7] Dynasim AB (1998), *Dymola - Dynamic Modeling Laboratory*, http://www.dynasim.se

[8] Buur, J. (1990), *A theoretical approach to mechatronic design*, PhD thesis, Institute for Engineering Design, Technical University of Denmark, Lyngby, Denmark.

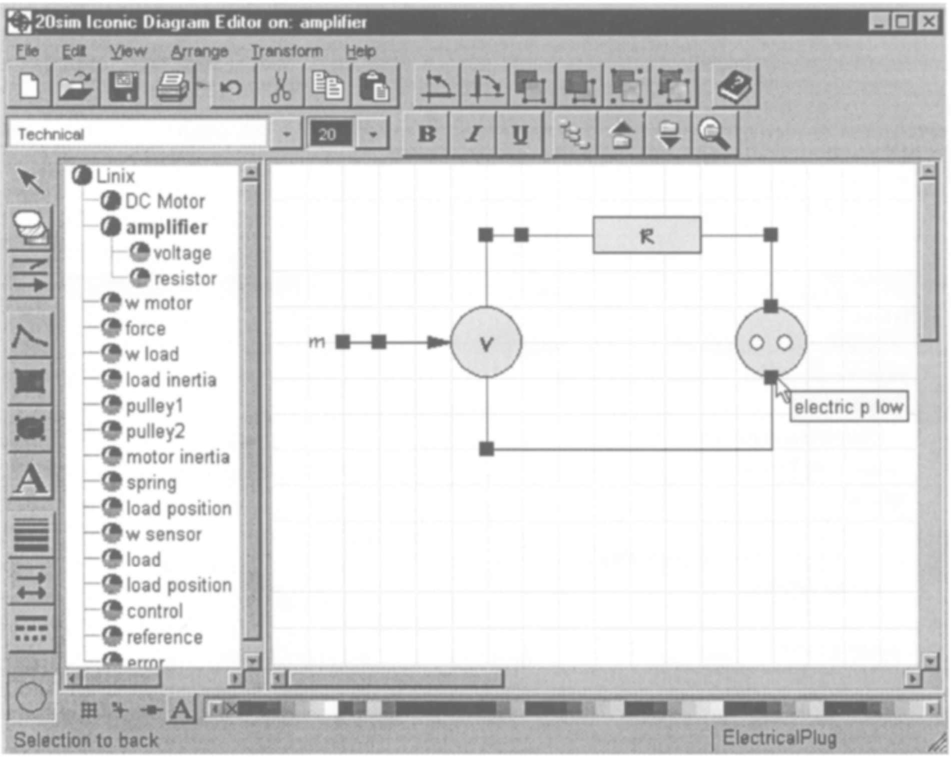

20sim Iconic Diagram Editor on: amplifier
File
Edit
View
Arrange
Transform
Help
Technical
20
B
I
U
Linix
DC Motor
amplifier
voltage
resistor
w motor
force
w load
load inertia
pulley1
pulley2
motor inertia
spring
load position
w sensor
load
load position
control
reference
error
R
m
V
electric p low
Selection to back
ElectricalPlug

Figure 9: The iconic diagram editor showing the amplifier implementation

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Integrated Computer Aided Programming (CAP) System for robots in reverse engineering applications

P. Leonov

Department of Manufacture Engineering, Technical University of Budapest,
E ép., 1 Egri József ut., Budapest 1111, Hungary

This paper presents a structure of an integrated computer aided programming system based on the robot use in reverse engineering applications. In a great number of reverse engineering tasks the use of a robotized system instead of combination of expensive CNC machines and even more expensive Coordinate Measuring Machines (CMM) can be extremely profitable. Questions of programming and realization of such systems are discussed.

1. INTRODUCTION

Reverse engineering is a procedure of the CAD model reconstructing from an existing prototype and then its following manufacturing. In recent years a wide variety of works has been published for the various aspects involved in the reverse engineering process. A unified review of the existing publications on this topic was published by Puntambekar [1]. Generally, different tasks including in the reverse engineering process are applied at small and medium sized companies. As it is well known, such companies are very important for the productivity in various mainly smaller countries. Small companies are easier in the position to produce special products according to the demand of the market in small lots due to their flexibility. Therefore, production automation is absolutely necessary for them. However there is a dark side of automation for such companies. This is a question of the cost of production tools and equipment. In fact, in literature the reverse engineering process in the frame of CIM-conception is usually discussed as a subject of the manufacturing by expensive CNC machine tools and a measurement by extremely expensive Coordinate Measuring Machines (CMM). In this situation the use of robotized systems which are much cheaper can be very lucrative. Furthermore, as it will be shown below, all the operations included into the reverse engineering process may be integrated into the same robotized system.

A lot of applications do not require an exact copy of the part, but rather the capture of ideas the designer originally had in mind when the part was being designed The implementation of inspection in combination with manufacturing by industrial robots provides an opportunity to solve a problem of reverse engineering in such cases much cheaper than it is in today's industrial practice.

The objective of this paper is to develop an integrated computer aided programming (CAP) system for robots in reverse engineer applications. In this work the author suggests a common structure of the computer aided programming of robots for wide scope manufacturing operations. The structure follows from the reverse engineering method used to be applied at the CNC manufacturing. Sequences of programming and flows of information and data are outlined.

In the case of reverse engineering manufacturing we use contact operations. Hence the force acting to the robot's end-effector should be taken into account, as the robot accuracy is less than that one for machine tools. In Podurajev, Leonov [2] a new methodical approach is discussed to construct an adaptive algorithm for contour following force control in deburring operations. This approach is based on the representation of robot dynamics in Riemann space.

In robotized manufacturing we must use small feed because of the robot's flexibility. Of course, the small feed limits the material removal rate and, hence, the productivity of process. So, in order to achieve the high speed of working tool and high productivity, the method of time-optimal processing should be used. Somlo, Podurajev [3] proposed a new approach to the solution of the trajectory planning problem for robots.

Accuracy is one of the most complicated obstacles for introducing robots in the Reverse Engineering Manufacturing (REM), because the robot accuracy is lower than that one for machine tools. It must be noticed that most of the developed method of robot calibration can not be implemented in REM due to specificity of performed tasks (for example, combination of measurement inspection and manufacturing into the same robotized system).

2. BASIC PRINCIPLES OF REVERSE ENGINEERING MANUFACTURING

In general, production process in modern manufacturing systems with robots or CNC-tools use consists of the following steps: designing a model, manufacturing and then part inspection. However, on many occasions one may be faced with a situations in which a part has to be manufactured based on existing prototype with no existing drawing (e.g. parts without an analytical description of their surface, products designed by artists, old and exclusive parts and etc.). In these cases, the CAD model has to be reconstructed from an existing prototype and then subsequently manufactured. The areas mentioned above are just some main ones in which reverse engineering can be successfully applied.

2.1. Reverse engineering process

Reverse engineering is a methodology for constructing CAD models of physical parts whose drawings are not available by digitizing an existing prototype, creating a computer model and then using its to manufacture the component. Although the reverse engineering process may seem to be the opposite of the conventional manufacturing process (start with CAD model and produce a part), the overall concept is the same. The main difference is that the existing prototype in reverse engineering embodies the product specification in conventional manufacturing. Another difference is that the generation of the CAD model from a design concept is an explicit process, while generation of CAD model from digitized data is an interferential process.

There are two main cases which reverse engineering can be apply in:

- **Object Reconstruction.** The prototype exists only in a few copies and it seems difficult or fully impossible to draw a CAD model because of its complication.
- **Measurement Inspection.** The prototype has just been manufactured by a machine tool or robot and the tolerances on the dimensions have to be checked.

There are a lot of examples of these cases: New Design, Old Parts Redesign, Worn or Damaged Parts, Replicating Components Acquired in War, Samples Inspection.

All the above mentioned processes are discussed as a subject of the manufacturing by CNC tools and measurement by CMM.

2.2. Production Flowchart of Integrated Process

The main idea of this paper is to apply robots at all steps of reverse engineering, direct manufacturing and parts inspection.

An integrated scheme involved these three stages is shown in *Fig. 1* in three vertical columns. This very detailed scheme introduces the sequence to produce a desired part from its prototype. A chain of actions in each stages is provided below.

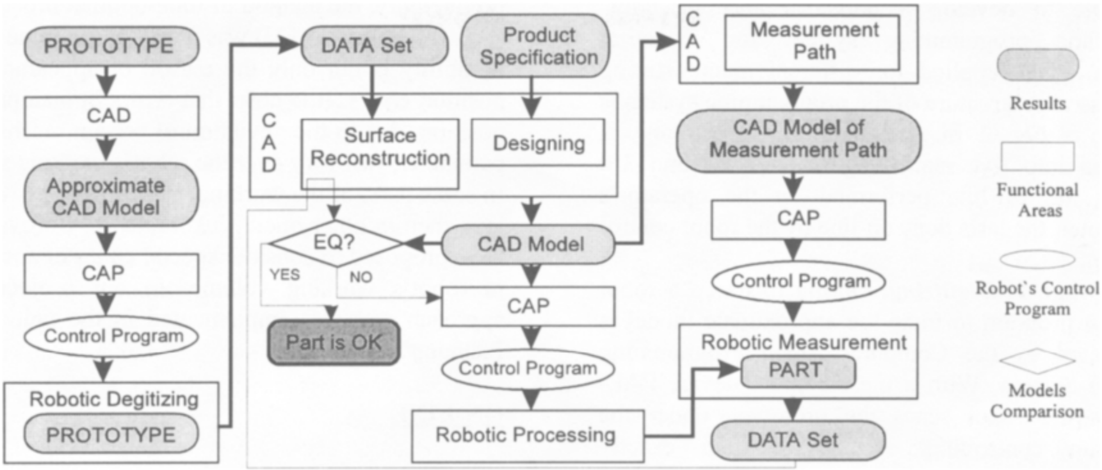

PROTOTYPE
CAD
Approximate CAD Model
CAP
Control Program
Robotic Degitizing
PROTOTYPE
DATA Set
Product Specification
CAD
Surface Reconstruction
Designing
EQ?
YES
NO
CAD Model
CAP
Part is OK
Control Program
Robotic Processing
CAD
Measurement Path
CAD Model of Measurement Path
CAP
Control Program
Robotic Measurement
PART
DATA Set
Results
Functional Areas
Robot's Control Program
Models Comparison

Fig.1. Integrated robotized manufacturing system structure and process sequences.

Column 1. Digitizing. At the CAD step a rough CAD model of the prototype part is designed by user. This model consists of only simple drawing entities like lines, arcs, circles and approximately describes the part's surface.

A robot control program to trace the approximate model is generated by the Computer Aided Programming (CAP) System. The robot equipped with a measurement device scans the prototype under program controlling. A set of data exactly characterized the surface is worked out by the control computer.

Column 2. Manufacturing. In case of reverse engineering surface reconstruction procedures are performed based on the data set from the previous step. If the product specification exactly determined, a drawing of the product is designed. In both cases a CAD model of the product or prototype is constructed. After CAP a robot program for workpiece processing is generated. Then, the robot cuts out the finished part.

If no measurement needs the production process is completed here.

Column 3. Part Inspection. To be sure that the given part is desired the measurement procedure can be performed. This procedure is the same as digitizing the prototype except the CAD model of measurement path, which is turned out from the exact CAD model on the previous stage. This allows fulfilling the measurement more preciously. A data set characterized the tolerances on the dimensions after surface reconstruction is compared with the model to improve the control program. If there are any deviations between the model and part the processing can be repeated till the desired result be approached.

The Computer Aided Programming procedure is practically the same in every sequence and described below with a robotics manufacturing system to be involved in the project.

3. AUTOMATIC ROBOT PROGRAMMING AND CONTROL

The robotized manufacturing system used for the first implementation consists of: a PUMA-560 robot, a control system SFERA-36, a personal computer with a data processing unit and the robot controller, a special table for fixing workpieces and force measurement, a pneumatic driver with a tool and the force/torque sensor, a workpiece for manufacturing or measurement. This complex may be used for performing the contact manufacturing operations as well as non-contact ones and measurement. In each case the robot's end-effector should to be equipped with special devises or sensors.

A simple and rapid way of robot programming is to apply automatic programming in which a robot program is automatically generated from the graphical description of a workpiese and some set of parameters. In presented work the author has aimed at using already existing software as much as possible to develop a universal and "friendly" automatic programming system for industrial robots, to be applied for different manufacturing processes. A structure of the programming system is shown in *Fig. 2.* In general, a program strategy is divided into five main stages. The 2nd and 3rd stages are off-line performed on the operator's computer, the lasts done on-line by the robot control system.

Stage 1. *Digitizing.* During this step a robot control program to trace the approximate model is generated by the Computer Aided Programming (CAP) System. With a measurement device PAL-Optics the robot scans the prototype under the program controlling. A set of data exactly characterized the surface is worked out by the control computer.

OFF-LINE

Stage 2. *Computer Aided Design.* On this step a full geometrical description of the workpiece is to be made by a CAD subsystem (e.g. AutoCAD). Other CAD systems providing common data file interface (e.g. DXF and IGES data file specifications) may also be applied.

Stage 3. *Computer Aided Manufacturing.* The file containing the geometrical description of the workpiece is accepted by the MASTERCAM package. MASTERCAM is a powerful system for programming CNC tools. The robot is considered as a particular CNC milling machine with five and more axis. The operator is ordered to determine tools and materials for processing and to choose a manufacturing operation. Then, a toolpath is automatically generated. The toolpath is stored in an NCI-type file in ordinary ASCII code. On this step a tool movement simulation may be also performed for the visual inspection and eventual correction of chosen parameters.

It is very important that the unit "Path Planning" (*see Fig. 3*) generates trajectory considering not only technological conditions but also accuracy and manufacturing productivity. In robotized manufacturing we must use small feeds because of the robot's flexibility which is one of the main reasons of low accuracy. Of course, the small feed limits the material removal rate and, hence, the productivity of process. That is why in order to achieve the high speed of working tool and high productivity, the method of time-optimal processing (Somlo, Podurajev [3]) was used. At the same time, flexibility is not only the reason of appearance of position errors. It is clear that errors in manipulator position, due to the deviation of certain mechanism parameter, will vary over the robot's working space. In some parts of the working space this error will be less, then in other ones. Vukobratovic, Borovac [4] have proposed to consider special zones of accuracy in robot's working volume. In our project this approach was also implemented in the unit "Path Planning".

ON-LINE

Stage 4. *Interface between computer and control system.* On this step a special postprocessor program processes the toolpath stored in the NCI-file. The postprocessor converts the file into appropriate commands of the robot programming language. For different control systems a set of postprocessors can be created to take into account specific characteristics each other. As a result, a robot control program is generated. Then, the control program is transmitted to the robot control system by a connection program through RS-232 serial port.

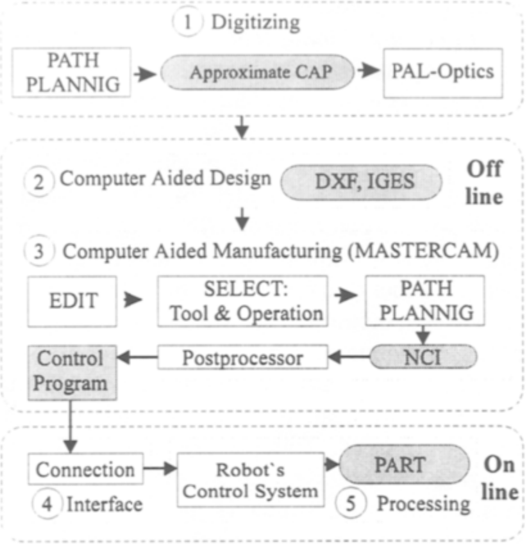


Fig.2. Steps of Computer Aided Programming.

Stage 5. ***Processing.*** The control program obtained at the last step is stored in the memory of the control system. Only a few characteristic points (generally just one) have to be located manually by the teach pendant. Then, the control program can be run under the adaptive force control algorithm described below.

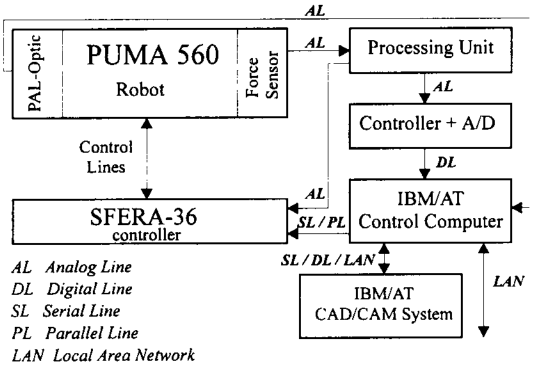

AL
PAL-Optic
PUMA 560
Robot
Force Sensor
AL
Processing Unit
AL
Controller + A/D
DL
Control Lines
AL
IBM/AT Control Computer
SFERA-36 controller
SL / PL
SL / DL / LAN
IBM/AT CAD/CAM System
LAN
AL Analog Line
DL Digital Line
SL Serial Line
PL Parallel Line
LAN Local Area Network

Fig.3. Communication diagram.

During the machining, some signals characterized the actual force come from the force sensor to the computer through the processing unit and A/D converter as shown in *Fig. 3*. A special computer program can use these values for robot controlling. An IBM/AT computer with CAD/CAM packages manages the overall system in off-line mode independently on other parts of the system and connects with one by serial/parallel interfaces or a local area network line.

CONCLUSIONS

A system for automatic programming and control of the described robotized manufacturing system has been presented. User without special knowledge in CAM or CNC programming should be in the position to operate the robot programming gaining high availability and product quality. In this way, the user has powerful software for optimal use of the system with a high degree of flexibility for the production. This makes it profitable to apply this system in companies with small batch sizes.

ACKNOWLEDGMENT

The author are indebted to Professor János Somló (Technical University of Budapest) and Professor Jury Podurajev (Moscow State University of Technology "STANKIN") for their leadership and valuable suggestion.

REFERENCES

1. Puntambekar N.V., Jablokow A.G. and Sommer H.J. III. “Unified review of 3D model generation for reverse engineering” in *Computer Integrated Manufacturing Systems*, 1994, 7(4), pp.259-268.
2. Podurajev J.V., Leonov P.V. “Adaptive force control and computer aided programming system for PUMA manipulators” in *Proceedings of the 4th International Workshop on RAA'95*, 1995, July 6-8, Portschach, Austria.
3. Somló J., Podurajev J. “Optimal cruising trajectory planning for robots” in *Mechatronics*, 1994, Vol.4, No.5, pp.517.
4. Vukobratovic M., Borovac B. “Accuracy of the Robot Possitioning due to their Kinematic Inaccuracies”, in *Mechanism and Machine theory*, 1993.

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
© 1998 Elsevier Science Ltd. All rights reserved.

Sensor controlled robotic arc welding integrated with simulation software

Per Hedenborn, Magnus Olsson and Gunnar Bolmsjö

Department of Production and Materials Engineering, Lund University
P.O. Box 118, 221 00 Lund Sweden

Abstract. This paper presents a totally sensory controlled system for robotic arc welding incorporating three-dimensional seam tracking with angular corrections. The implementation of real time process models is also described which allow corrective actions to be taken in real time on the basis of the actual features of the workpiece. Information from the sensor system can be fed back to a supervisory system consisting of simulation software for robotic applications, adding possibilities for on-line calibration of the workcell. The concept of a virtual sensor model is described where new possibilities are added during off-line simulation of the system. This includes a better representation of the real workpiece, giving the operator a tool for studying how variations in the geometrical features of the welding joint affects the behaviour of the system when the sensor data is processed by adaptive, real time process models. Models, that are implemented in the simulation software allowing the user to test and verify the behaviour and response of the total system in an off-line environment.

1 Introduction

Demand for higher productivity in combination with more flexible production, down to one-off production, makes it necessary to minimise the programming time as well as ensuring that the required quality level will be met. The introduction of sensor controlled robots together with off-line programming capabilities of the robot is an attractive alternative. The need for explicit programming of the trajectory is reduced by the approach of offline programming and sensors capable of guiding the robot relative the work piece, detecting local deviations.

The possible applications for the presented system range from robotic assembly to arc welding. In the latter case, the process itself is quite sensitive to variations in joint alignment and geometry. In general, the process models developed maps the observable parameters to the controllable [2]. This allows for taking corrective actions based on the actual features of the joint. One problem with this is that the function issuing these actions has no information about the kinematics of the robot or its environment. By integrating sensor data with the geometrical world model data and kinematics data related to robotic motions a lot can be gained.

Simulation and the generation of robot programs in an off-line environment always deal with uncertainties. The workpiece might e.g. not be fitted correctly, causing undesired results. Even if sensor information is fed back, unwanted interruptions might occur. As an effort to minimise these interrupts, a virtual sensor can be created where uncertainties about the position, joint features etc. is added. The response from the sensor system and the associated process models are simulated giving the operator an additional tool in the process of generating a program off-line. Another important aspect is the ability to test and verify software in embedded systems [3].

The approach is suitable both for optimising production of existing products where improved control strategies can be tested and when new products are to be introduced. Programming of the robot system and control of the process is aided by creating models that corresponds to the actual workpiece.

2 Robotic System

The robot system used within the project is based on work made at the Department of Automatic Control, Lund University, where an open controller has been developed. The controller has a modular structure allowing the implementation of external control at different levels. In addition, the software is built around dynamically linked objects that can be introduced or replaced during run-time [6]. Teleoperated robotic applications have already been implemented with possibilities of controlling the robot externally from graphical simulation software. Communication routines for reporting the current position of the robot to a supervisory system are also available.

Normally, robot controllers lack more advanced general-purpose interfaces capable of receiving process parameters together with positional information. If they exist, they are usually dedicated to a specific type of sensor. The approach within the project is to introduce a specific sensor board capable of processing sensor information and any information associated with the trajectory. This information is then passed on to the robot controller. The flexibility of the controller makes it possible to implement the necessary software for a wide variety of applications.

3 Sensor system

Robotic arc welding relies on good joint preparation, accurate positioning of the workpiece and a minimum of distortions due to the heat input. Any deviation from the nominal values or in the actual position will result in poor quality. Welding of large structures or in short series and one-off production presents a problem since it requires flexible fixtures and clamping devices together with time consuming programming of the workcell. In these cases, the use of seam tracking sensors is an attractive option since it is capable of identifying and following the seam and this requires, at least in theory, less programming of the robot. Furthermore, the need for clamping devices and fixtures is reduced.

A sensor system consists of a camera mounted in front of the torch together with a controller for processing of the video signal and the two dimensional image of the joint, see figure 3:1. The mounting of the camera in front of the torch allows the system to identify, measure and determine the position of the joint relative the torch in advance. The information is stored in a buffer and used for generating a trajectory.

The information obtained by the sensor is processed in the sensor controller where the *(y,z)* coordinates representing the profile of the joint, is reduced to line segments. The break points of these segments are then used for identifying the type of joint and other features such as, included angles, mismatch, area, etc. The nominal target position is calculated on the basis of the break points, see figure 3:2.

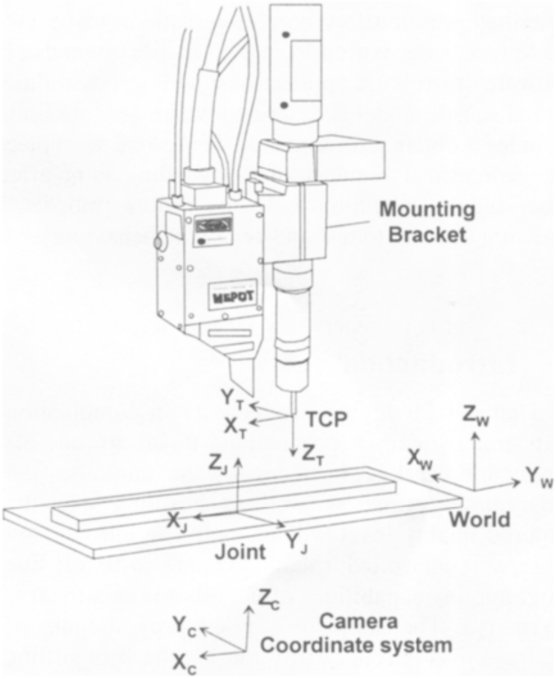
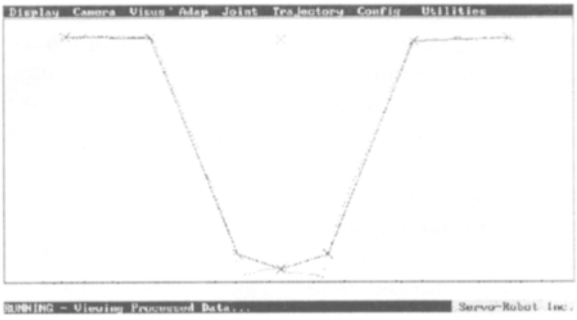


Figure 3:1 *Camera mounted in front of the welding torch and the different coordinate systems.*



RUNNING - Viewing Processed Data...

Figure 3:2 *Typical image of a V-groove including the breakpoints and calculated target position.*

In general, poor sensor interfaces of robot controllers' limits the implementation of more advanced seam tracking, i.e. three-dimensional seam tracking in combination with real time process models. However, in this case the presence of an open control structure allows the implementation of several important features of the system. A sensor board is added to the controller co-existing on the same bus as the controller. Information from the sensor system is transferred to the sensor board for further processing. This includes the generation of a trajectory with its associated parameters such as speed and orientational corrections. The robot is in this mode fully sensor driven and no actual program exists and the seam tracking sensor guides the robot as long as the correct joint is visible or an end condition is reached.

Typical end conditions include, end of seam, signal from other sensors that an obstacle has been detected or from a supervisory system, preventing collisions between the structure of the robot and the workpiece.

In addition to generating a trajectory in real time, the actual geometrical features of the joint is extracted from the sensor system. This information is used as an input to real time process models, which adjusts the position of the torch relative the workpiece based on knowledge about the behaviour of the process.

The problem with adaptive welding is that the observable parameters differ from the controllable. The process models developed maps the observable parameters to the controllable [2]. A typical example is the fact that the sensor is able to detect and measure the gap size of a fillet joint but unable to control it since this is a result of the joint preparation and the setup. Using a real-time process model, the system is able to adjust controllable parameters such as welding speed, conctact-tube-to-workpiece distance, orientation relative the workpiece and adding offsets, ensuring both quality and productivity. Additional strategies can be implemented, such as, weaving motions with various patterns if the gap increases, preventing burn-through during root pass welding. Parameters in the welding power source can also be updated based on sensor information about the actual geometrical features of the joint in each segment.

4 Simulation software

For graphical simulation, IGRIP[1] is used allowing simulation of robotic tasks with real time graphics. It is also possible to simulate the dynamic behaviour of the mechanical structure, create obstacle free trajectories, create and update the geometrical model of the workcell. The flexibility of the software allows for implementation of custom control programs of devices and manipulators. Furthermore, information can easily be transferred between external programs and IGRIP.

For programming devices in a workcell, a high level language GSL, Graphical Simulation Language, is used. This is a flexible language that can be used to define sub-tasks in a model. The model of the robot workcell can be updated to reflect the real world by sensor information or when a sub-task is executed.

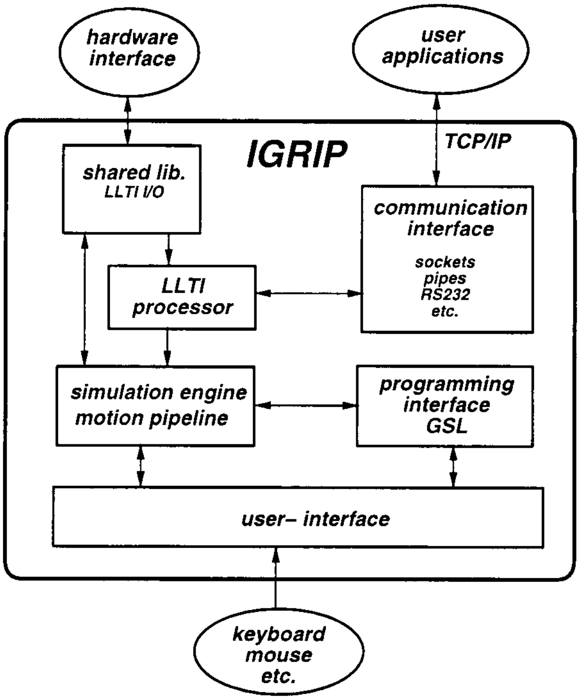


Figure 4:1 *Modules in IGRIP.*

An extension to IGRIP called LLTI, Low Level Tele-robotics Interface, can be used for both extracting and changing all types of data in the world model. With LLTI, a user-defined procedure is executed approximately 30 times per second enabling the exchange of information with other processes in the computer. External equipment can be integrated in the same way by the use of other interfaces such as serial port or TCP/IP communication, see figure 4:1.

---

[1] IGRIP is a trademark of Deneb Robotics Inc.

In figure 4:2, a basic integration between the simulation software, IGRIP, and a robot controller is shown. The robot and its environment are modelled within the simulation software giving a representation of the geometry of the workcell as well as the kinematic behaviour of the robot.

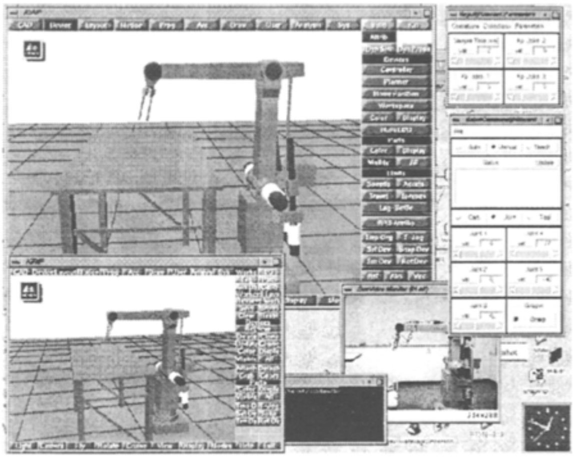


Figure 4:2 *Screen capture of tele-operated robot with visual and positional feedback [1].*

In this experimental setup, IGRIP is used to generate trajectory information to the robot. During the simulation of the robot, the LLTI interface is used to extract the current state of the virtual robot at constant time intervals. This information is passed on to the robot controller through TCP/IP over the Internet. The robot controller is located at another department at the university some 100 meters away. In addition, the robot controller sends back the current position to IGRIP, adding positional feedback.

This feature alone adds little or nothing to the system compared to traditional off-line programming where the program created in the simulation is converted to a static program and then downloaded and executed.

An improvement is if visual feedback is added, so that the operator of the simulation is able to see the execution of actual robot. The operator could then execute sub-tasks and on the basis of the outcome, initiate additional sub-programs. This is more of a traditional telerobotic system.

In both the off-line and telerobotic case, there is no or very limited information about if the geometry created in the simulation software is in accordance with the real objects. Positional and orientational deviations can be detected by the use of local sensors and the trajectory adjusted accordingly. The problem is that the sensor system usually has no information about the world model or the current kinematic solution of the robot. This means that the sensor system can guide the robot in such way that collisions occur or towards singularities. By using the capabilities of the simulation software together with positional feedback and information about local features of the workpiece, new possibilities arise.

The information from local sensor can be used for updating the geometrical world model so that it matches the virtual model. In addition, the detection of any deviations from the original world model allows the system to verify that no collisions will occur, adding possibilities for re-planning of the trajectory. Furthermore, the new situation might require a new configuration to be selected for the robot in order to avoid singularities and/or out of workspace conditions.

The application used for the experimental study is robotic arc welding and this is a very interesting application since the process is sensitive to local geometrical variations as well as it relies on constant relation between the torch and the workpiece. The complexity of the process can be modelled allowing reactive actions to be taken based on the joint geometry, constraints of the robot and the environment. The integration of sensor data to the simulation system opens up new possibilities both for logging positional and process data, indicating to the operator where possible weld defects might occur and for re-planning operations.

An example of re-planning operations includes obstacle avoidance based on sensor information and the selection of a more suitable configuration of the robot, given the new circumstances detected by the sensors.

5 Virtual sensor model

Taking things a step further, the creation of a virtual sensor that uses nominal data in simulated scenarios results in a more refined model of the workcell. A virtual sensor model has been introduced to emulate the behaviour of the physical sensor device. Today the user of robot simulation systems is limited to a static world model. This model can not in a proper way describe a world with its uncertainties and the sensor used to detect these. The virtual sensor model can be used in the implementation phase of a new system or process, to improve the result.

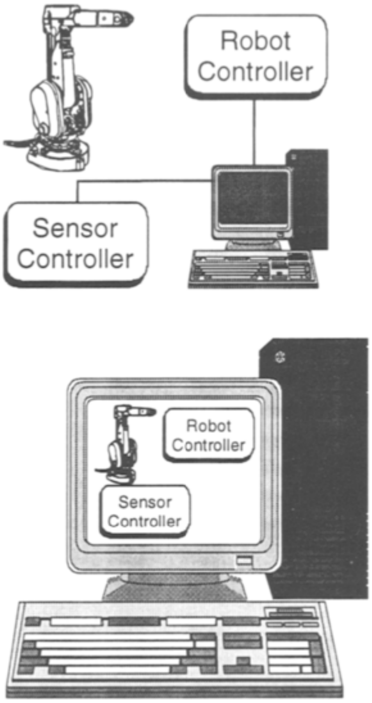

Robot
Controller
Sensor
Controller
Robot
Controller
Sensor
Controller

Figure 5:1 *The real system structure vs. the virtual scenario.*

In figure 5:1, the difference between the real structure of the system and the virtual modelling of the same, is shown.

The virtual sensor model gives the user of the simulation system a new interface to influence the simulation. In the presented case both process parameters like gap and plate-angles can be changed during a simulation, see figure 5:2. The process model will then take actions on the basis of this simulated sensor input, and change the robot movement according to the output from the process models. This includes the start of a weaving movement if e.g. the gap size exceeds a given threshold, increase the stick-out or change the orientation of the welding torch. The new robot movement can immediately be studied in the simulation software.

It is also possible to study how the position of the workpiece affects the behaviour of the robot by introducing vertical and horizontal offsets and orientational variations. These changes can cause unforeseen results like singularity problem of the robot arm, exceeding joint-limits etc.

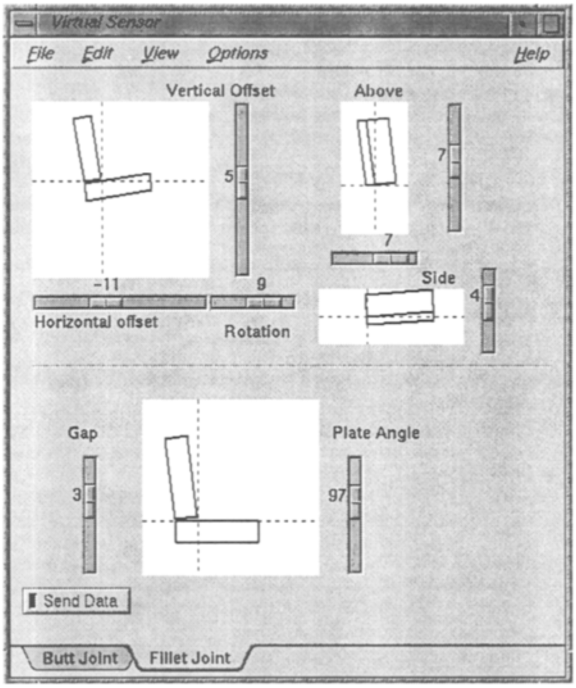

Virtual Sensor
File
Edit
View
Options
Help
Vertical Offset
5
Above
7
7
-11
9
Side
4
Horizontal offset
Rotation
Gap
3
Plate Angle
97
Send Data
Butt Joint
Fillet Joint

Figure 5:2 *Example of the user interface for the virtual sensor.*

To be able to fully understand the behaviour of a complex system a static model is too limited and the introduction of a virtual sensor model provides an important tool to the operator. A tool that allows the operator to verify that the program generated is capable of handling variations that lies within the specified tolerances.

6 Summary

The presented approach of a system capable of adapting to changes in the real world as well as simulating them, gives the operator a new set of programming tools. The programming of a robotic cell moves towards a task oriented approach. The operator of the off-line simulation software package is supported by process knowledge embedded in real-time process models with possibilities of simulating geometrical variations. The approach gives a more reliable result since the effect of different variations can be tested.

The robot program is focused on the higher level, i.e. welding sequence, joint type and avoidance of known objects. The sensors then guide and control the robot locally, ensuring productivity and quality by adapting to any geometrical deviation. In addition, by performing the programming off-line the actual robot is able to stay in production while programs for new products is generated.

Creating virtual sensor models that interacts with process models mapping observable parameters to those that are controllable also allow extensive testing and verification of control strategies and algorithms. By having the same models in the simulated environment as those implemented for real-time production, the correspondence between the actual system and the virtual is ensured.

7 Acknowledgements

Parts of the work presented in this paper are sponsored by the Swedish National Board for Industrial and Technical Development (NUTEK) within the framework for Mobile Autonomous Systems (RAS). Part of the work is a collaborative effort by the Department of Automatic Control, Department of Industrial Electrical Engineering and Automation and the Department of Production and Materials Engineering at Lund University, Sweden.

References

1. Brink K., Olsson M., Bomsjö G.: Increased Autonomy in Industrial Robotic Systems: A Framework, *Journal of Intelligent and Robotic Systems* 19, pp 357-373, 1997.
2. Hedenborn P., Bolmsjö G.: Task Control in Robotic Arc Welding by Use of Sensors, In: *Proc. International Conference on Machine Automation. ICMA '94*: Tampere, Finland, 1994.
3. Ju-Yeon Jo et al.: A New Role of Graphical Simulation: Software Testing: CAISR technical report TR-96-115, Cleveland Ohio, United States, 1996.
4. Madsen Ole: Sensor Based Robotic Multi-Pass Welding, PhD thesis, Department of Production, Aalborg University, Aalborg , Denmark, 1992.
5. Nayak N. and Ray A.: Intelligent Seam Tracking for Robotic Welding, Springer-Verlag, 1993.
6. Nilsson Klas: Industrial Robot Programming, PhD thesis, Department of Automatic Control, Lund Institute of Technology, Lund, Sweden 1996.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Simulation of Continuous Process Variations in Production System Design

Jan Oscarsson[a] and Philip R. Moore[b]

[a]Centre for Intelligent Automation
University of Skövde
Sweden

[b]Mechatronics research group
De Montfort University
UK

With reduced time available and demands for cost reduction computer aided robotics (CAR) may be used in industry more frequently for a variety of tasks. Normally CAR models only represents nominal relations which results in a deviation between real and virtual world. One type of deviation is continuous process variations, which ought to be represented in CAR models. This paper defines the term continuous process variations and suggests a model for implementation in CAR. Typical examples of continuous process variations are robot repeatability, varying part shapes due to manufacturing tolerances and sensor repeatability. In an experiment effects on a assembly process are studied.

1. Introduction

Due to demands for shorter design-to-market lead times and reduced product and production system life cycles expose the process of developing and programming manufacturing systems. There is a need for a more rapid method for design of new manufacturing systems. Use of prototypes and experimental set-ups for tests and evaluations of various manufacturing alternatives need to be reduced. The rapid development of both hardware and software in conjunction with improved usability suggests an extended use of software tools as a possible alternative. Use of computer tools has increased in importance for the task of designing and developing products and production systems and now offers some significant advantages compared to conventional methods [1].

One computer tool possible to use for the design of manufacturing systems such as industrial robot based assembly system are computer aided robotics (CAR). CAR are graphical computer systems originally intended for simulation and off-line programming of robots. CAR systems are widely used in industry for tasks like simulation and off-line programming of robots. In CAR robots are modelled using 3D CAD objects representing the body of the robot along with implemented kinematic models (including inverse kinematics), robot dynamics and control systems [2]. Properties like mass, inertia etc. may also be modelled. From libraries selection of basic components such as robots, tools, sensors etc. may be done. In CAR its possible to simulate not only the manufacturing device, by inclusion of models of parts, feeders, tools, fixtures and other peripheral equipment the complete manufacturing system may be simulated and analysed. The fact that the product is included within the simulation model means that not only demands from the product on the manufacturing system can be accounted for, but also constraints on the product imposed by the manufacturing system can be observed, see figure 1.

The task of manually examining and analysing effects from complex interactions between the product and the manufacturing system may be difficult or even impossible. Implementation of advanced analysis and report functions in CAR significantly improves its usefulness.

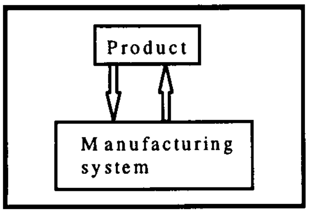


Figure 1: Interaction between product and the manufacturing system.

Even though use of CAR offers several advantages and opportunities, a number of limitations and restrictions exists. One problem is the fact that several types of deviations exist between the virtual and the real world, an example of such a deviation is continuous process variations.

2. Continuous process variations

Continuous process variations (CPV's) concerns those process variables within a manufacturing system whose variation influences or affects the manufacturing process. CPV can be defined as:

"A continuous variation of a process variable in a manufacturing system which displays a stochastic variation without being influenced by discrete events in- or outside the manufacturing system."

From the definition, one significant type of CPV easily can be identified, namely manufacturing tolerances. However, the definition of CPV covers a broader area which includes a large number of different variation types. CPV´s can either be of a general or more application specific type. General CPV´s are always to some extent present in a manufacturing system, affecting the production process more or less in terms of quality, productivity and availability. Examples of general CPV´s in a robot based assembly application are variations in object locations, part shapes, robot repeatability and sensor repeatability. Application specific CPV´s are dependent on the manufacturing task performed and the current process.

2.1. Representation of CPV's

Representation of CPV's in CAR requires a model to be implemented in the virtual environment. Such a model may be developed using the reference model shown in figure 2.

The reference model is divided into two distinct elements. The first element, variation, includes representation of the physical characteristics and the variation behaviour. This is explained by the following example:

On a blueprint two holes with the same diameter are drawn. Attached to their diameter dimension is a tolerance value, defining the tolerance zone, namely ± 0.1 and ± 0.5 mm for each hole respectively.

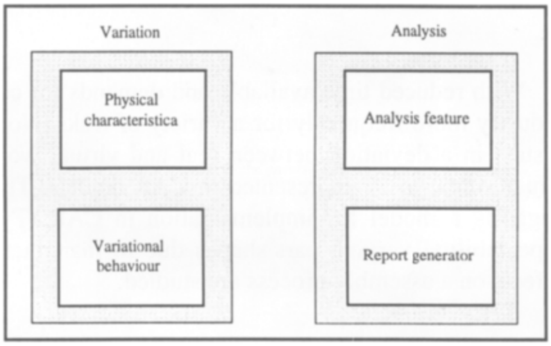


Figure 2: Continuous process variation reference model.

In this example the physical characteristics, varying hole diameter, are identical for both CPV's. However, their variational behaviour parameter, namely the tolerance zone for the holes differ.

Modelling of the physical characteristics has to be defined and implemented once for both holes, or generally once for each type of CPV. However, the variational behaviour parameters has to be defined individually for each specific CPV. When assigning parameters for representation of a specific behaviour it is critical that they corresponds closely to the variation behaviour in reality. Data regarding real variation behaviour can be obtained either from measurements or from design specifications related to each feature.

The second element in the reference model is an analysis function which is required in order to determine the impact from each CPV on the manufacturing process. For both reliability and feasibility reasons does the analyse function need to be automated. Useful is then a programmable collision detection function. With programmable means that one can define which objects and when to be checked for collisions.

The task for the report generator is to aquire and present relevant information during a simulation. Reports ought to be designed fitting different end users. As an example, purchasers of equipment need a device specification when personnel working with production quality are more interested in functionality and process reliability.

When performing simulation in CAR which incorporates CPV's those process steps shown in figure 3 can be used.

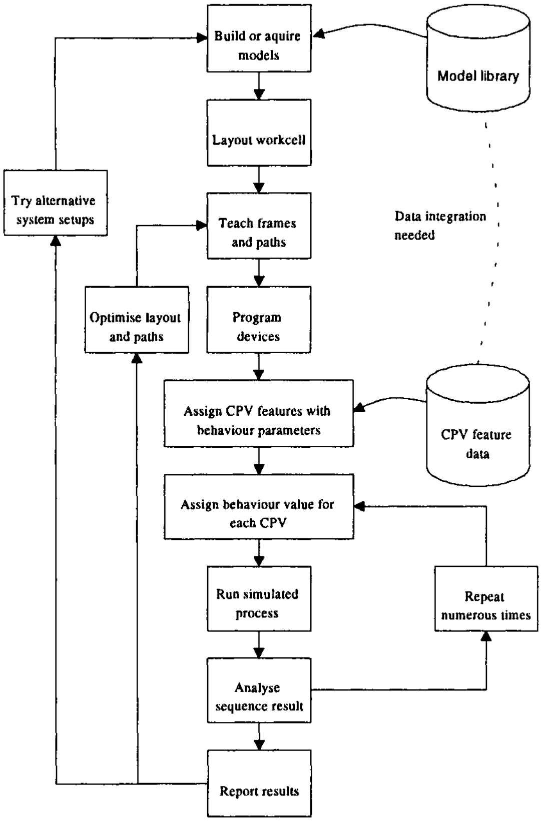

Build or aquire models
Model library
Layout workcell
Try alternative system setups
Teach frames and paths
Data integration needed
Optimise layout and paths
Program devices
Assign CPV features with behaviour parameters
CPV feature data
Assign behaviour value for each CPV
Run simulated process
Repeat numerous times
Analyse sequence result
Report results

Figure 3. CPV simulation process steps.

3. Simulation of CPV's

An ability to simulate CPV's and analyse their effects in a manufacturing process may be of great importance, especially if CPV's present in a manufacturing process results in a situation with complex relations and where combinatorial effects among CPV's are difficult or even impossible to model and predict using manual methods. Numerous repeated simulated sequences with altered combinations of input parameters offers a possibility to predict the overall system output.

3.1. Positioning repeatability

Positioning repeatability is one important parameter for a manipulator. During the design phase of a manufacturing system positioning repeatability has to be considered with respect to those demands the product puts on the system. An assembly operation will normally be quite error prone if parts fit-in tolerances far exceeds the repeatability. Positioning repeatability includes both positioning devices operating in joint space and devices operating in cartesisan space using kinematic models.

3.1.1. Repeatability in joint space

Simulation of repeatability in joint space addresses representation of joints which are controlled and positioned individually. Joint positions are defined as nominal goal positions. Additional parameters for a motion like speed or acceleration may be included when operating the joint to the goal position. In order to represent positioning repeatability a delta position value is calculated. The resulting goal position for the joint is then defined by adding the delta position to the nominal goal position:

$$\mathrm{Pos}_{\mathrm{resulting}} = \mathrm{Pos}_{\mathrm{nominal}} + \Delta\ \mathrm{Pos}$$

The delta position value is obtained using a randomising function whose operation input data is based on the CPV behaviour parameters. If several joints performs a co-ordinated motion repeatability are assigned to each joint. Simultaneous operation between independent or interpolating joints can then be simulated. By inclusion of repeatability parameters for each joint an overall system behaviour can be simulated and analysed.

3.1.2. Repeatability in cartesian space

Motion of manipulators in cartesian space is based on a kinematic model. Based on the kinematic model motions and poses defined in cartesian space are converted to motions and positions in joint space. During operation each joint of the manipulator are moved in order to make the TCP frame achieve a desired goal pose. Poses are defined in cartesian space by frames, each frame represented by a 4 x 4 transformation matrix (T).

Simulation and representation of robot repeatability may be done either by identifying parameters for each link in joint space or by manipulating the cartesian goal frame. As defined by the ISO 9283 "Manipulating industrial robots - Performance criteria and related test methods" standard, positioning repeatability for industrial robots is defined in cartesian space. It is therefore advantageous to simulate repeatability in cartesian space. This may be done by adding a delta frame to the nominal goal frame. In simulation this is done by assigning values to a delta matrix and then add that matrix to the nominal goal matrix obtaining a resulting pose matrix:

$$\mathrm{T}_{\mathrm{resulting}} = T_{\mathrm{nominal}} + \Delta\,\mathrm{T}$$

Values are added to each element in the delta matrix using a randomising function. This way may robot repeatability in cartesian space be implemented in simulation. Further more, by assigning different variation behaviour parameters repeatability may be simulated for each robot individually. Inclusion of repeatability in simulation offers an excellent opportunity to study the performance of a robot in a manufacturing process.

3.2. Location repeatability

In many manufacturing processes objects need to be fixtured. Fixturing of parts are related to repeatability issues, especially if several fixtures are circulating in a manufacturing system used as carriers of products. Their individual shape variation will then cause objects to occur in slightly different positions. Simulation of location repeatability may be done using the kinematic formulation described above, but with manipulating the fixtured part instead of the joint.

3.3. Varying part shapes

Varying part shapes are mainly caused by manufacturing tolerances and are perhaps one of the most common types of CPV's. Manufacturing tolerances are considered from two different perspectives. First with respect to functional and quality requirements of the product, and second with respect to the manufacturing system capability. Several methods for simulation of manufacturing tolerances has been reported [3]. Which method to be used is most of all an issue about how part geometry data are generated and captured. A kinematic formulation described by Rivest et al. and Whitney et al. [4,5] can be used for representation of manufacturing tolerances. This method is suitable when modelling manufacturing tolerances in a CAR system, mainly because most CAR systems do handle kinematic descriptions very well, but perhaps have a bit ordinary CAD-module in which varying part shapes due to manufacturing tolerances normally not are represented.

Geometric variations and tolerances are represented by using homogenous matrix transforms. Variations in 6 DOF can be established. Both relations between objects and relations between features in an object can be modelled. Nominal relations are modelled by individual matrixes and chains of relations are represented by chaining together the 4 x 4 transformations that relate successive parts. By adding a ΔT-transformation matrix, an adjustment transform, to each nominal transform tolerances between features can be described. Datums, or reference frames are the basis for relating features.

By adding a ΔT-transformation to the planar part dimensioned with L in figure 4, the lateral tolerance (t) can be described:

$$ {}^{0}_{1}T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & L \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

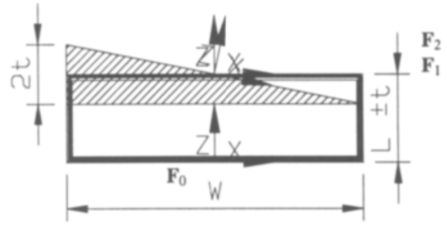


Figure 4. Lateral (t) and angular (θ_y) tolerances.

In a similar way angular tolerances can be described. A 3rd successive frame, $\mathbf{F}_2$, represents the angular error. The angels for θ_x and θ_y are obtained using small angle approximation ($W>>t$). Figure 4 shows an angular error (θ_y) in one dimension, rotation round Y:

$$ {}^1_2T = \begin{bmatrix} 1 & 0 & 2t/W & 0 \\ 0 & 1 & 0 & 0 \\ -2t/W & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} $$

4. Experiment

In an experiment three types of CPV's were implemented using the proposed reference model. In the experiment individual as well as combinatorial effects of CPV's were studied and analysed. An assembly operation was simulated in which a conventional industry robot performed a peg-in-hole mating operation, see figure 5.

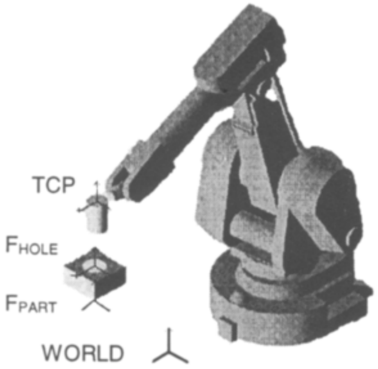


Figure 5: Simulated assembly operation

Nominal relations in the virtual assembly cell are:

$$ {}^{TCP}_{hole}T = \left[{}^{robot}_{TCP}T\right]^{-1} \bullet \left[{}^{world}_{robot}T\right]^{-1} \bullet \left[{}^{world}_{part}T\right] \bullet \left[{}^{part}_{hole}T\right] $$

To the assembly operation following CPV's were added:

1. Robot repeatability in cartesian space [TCP], 3 DOF.
2. Part manufacturing tolerance, hole position [F_{HOLE}] relative part corner [F_{PART}], 2 DOF.
3. Part location, [F_{PART}] relative world frame [WORLD], 1 DOF.

Given these CPV's, following relationships can be described:

$$ {}^{TCP}_{hole}T = \left[{}^{robot}_{TCP}T + \Delta T_{repeatability}\right]^{-1} \bullet \left[{}^{world}_{robot}T\right]^{-1} + \left[{}^{world}_{part}T + \Delta T_{location}\right] \bullet \left[{}^{part}_{hole}T + \Delta T_{tolerance}\right] $$

When the cell had been modelled and all devices programmed initial experiments were performed in which only one CPV feature at a time was simulated For each CPV several behaviour parameters were assigned, and for each parameter value the assembly sequence was repeated 100 times. A programmable collision detection function was used to analyse the result of the operation. Results from the initial simulations are shown in figure 6 and 7.

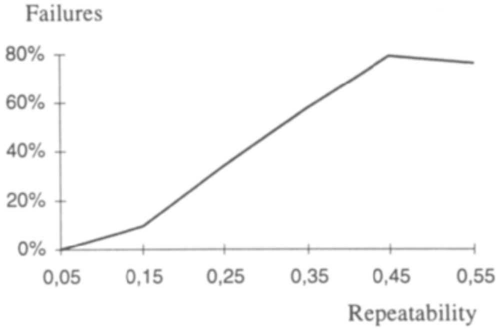


Figure 6. Results from simulation of robot repeatability.

As can be seen there is a relatively linear relation between the variation parameters and the frequency of failures. After the initial experiments combinatorial effects between the CPV's were studied. Combinations of different behaviour parameters were studied during repeated simulations. Variation behaviour parameters for each simulation are shown in figure 8, and results from the simulations are shown in figure 9. As can be seen, even relatively small variation intervals results in a relatively high frequency of failures.

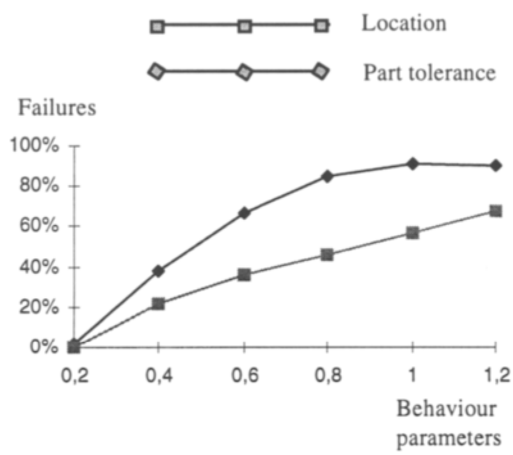


Figure 7. Results from simulation of part location respectively manufacturing tolerance.

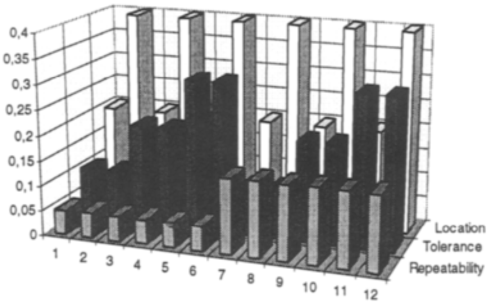


Figure 8. Variation behaviour parameters.

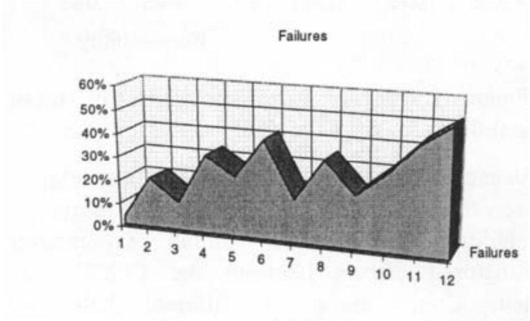


Figure 9. Results from simulated assembly sequence, failure frequencies from combinations of CPV's

5. Conclusions and future development

From the experiment one can conclude that it is possible to represent and simulate continuous process variations in CAR and study their effect on the manufacturing process. The proposed reference model for implementation may then be used. Continuous process variations may have a considerable effect a manufacturing process, hence need they to be considered when designing a manufacturing system. Use of CAR offers then a possibility to rapidly and reliably simulate and evaluate different alternatives, including different variation behaviours.

Future development in the area of simulation of continuous process variations includes several issues. Integration of data is a most desirable development for the end user, see figure 3. Ideally should CPV data be stored along with geometrical data in an extended CAD file, a product model. As an example should repeatability data be assigned to each virtual robot model. Another area in which simulation of CPV would be of interest is in the study of robot adaptiveness and error recovery strategies. The area has been subject for much work and use of CAR offers several advantages.

References

1. Kronreif, G. (1995), "ROMOBIL/SITAR - A user oriented simulation package for roboterized assembling," *Proc. of 26th Int. Symp on Industrial Robots*, Singapore.

2. Craig, J. J. (1988). "Issues in the Design of Off-Line Programming Systems". *Fourth International Symposium of Robotics Research*, MIT Press, USA.

3. Juster, N. P., (1992), "Modelling and representation of dimensions and tolerances: a survey," *Computer Aided Design*, vol. 24(1), pp. 3-17.

6. Rivest, L., C. Fortin and A. Desrochers, (1993), "Tolerance modelling for 3D analysis; Presenting a kinematic formulation," *Proc. of 3rd CIRP Seminars on Computer Aided Tolerancing*. France.

7. Whitney, D. E., O. L. Gilbert and M. Jastrzebski, (1994), "Representation of Geometric Variations Using Transforms for Statistical Tolerance Analysis in Assemblies." *Research in Engineering Design*, vol. 6, pp. 191-210.

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)

© 1998 Elsevier Science Ltd. All rights reserved.

Interfacing Autolev to Matlab and Simulink

M. Abderrahim[a*] and A.R. Whittaker[b]

[a]Departamento de Ing., Universidad Carlos III de Madrid,
Calle Butarque 15, 28911 Leganés, Madrid, Spain

[b]Department of Mechanical Engineering, University of Glasgow,
James Watt South Building, University Avenue, Glasgow G20 8QQ, UK.

The work presented in this article consists of a front and a back end to AUTOLEV, a commercial symbolic manipulation program designed to generate the equations of motion for multibody systems. The front end is a program which makes introducing the details of a serial link manipulator to AUTOLEV very easy and simple, even for individuals with very little knowledge of dynamics, whereas the back end consists of an interface program between the AUTOLEV generated program and MATLAB and SIMULINK. The interface program produces a MATLAB or SIMULINK 'm file' or 'mex file' ready to use with significantly improved simulation time. In addition, the use of the model with SIMULINK allows a larger choice of analysis and control design activities. The presented results and comparison of simulation time using different software illustrates the improvement obtained after using the suggested interface

1. INTRODUCTION

Over the last few years several modelling and simulation software packages of multibody dynamics have been developed [1]. Although these packages were developed for the same purpose, they differ in several aspects such as the algorithms they use, their efficiency and their ease of use. Between the several packages described in [1] only a few of them are designed specifically to deal with robotic manipulator dynamics. During the study which lead to writing this article three modelling and simulation software packages have been looked at closely for the purpose of evaluation and comparison to decide which of them would suit better the modelling of manipulators. SD/FAST, DADS and AUTOLEV are all good at producing models and simulation codes but provide little support for other modelling activities such as control design. From the ease of use point of view DADS is conceptually straightforward to use, however the fact that information and data have to be keyed in a certain format for each single element of the model makes it sometimes quite tedious. SD/FAST on the other hand is very easy and simple to use. The information is provided in a form of a descriptive file that can be written by almost anyone. The model is coded in FORTRAN and generated in the form of routines which require a simulation medium such as ACSL to be executed. AUTOLEV also provides the model in a symbolic form coded in FORTRAN but these are stand alone simulation programs and do not require separate simulation software. Unlike many other modelling packages such as SD/FAST, DADS and ADAMS which can be run by non-specialists in the matter of Dynamics, AUTOLEV can be used effectively only by individuals with good background in dynamics. The user has to be familiar in particular with the virtual forces method known also as Kane's Method. An initial evaluation of some modelling simulation software is summarised in [2].

A convenient modelling package would be as easy to use as SD/FAST and with a moderate cost such as that of AUTOLEV and with the possibility to connect with some control design and analysis programs such as MATLAB/SIMULINK. The following sections provide a description of a front and a back end written to complement AU-

---

[*]New improved versions of the Commercial packages SD/FAST, DADS and AUTOLEV have appeared on the market, since this work was completed.

TOLEV in order to achieve our idea of convenient modelling software.

2. AUTOLEV

AUTOLEV is a commercial advanced interactive symbolic manipulation program designed to assist the user in generating the equations of motion for multibody systems. It generates complete stand alone FORTRAN simulation code. AUTOLEV can be used effectively only by individuals who are familiar with the dynamics method known as Kane's formulation. The user must supply AUTOLEV with specific information for each element of the dynamical system under consideration. The information required by the software is provided in a given order and consists of the following:

1. An expression for the angular velocity in inertial space of each rigid body in the system.
2. An expression for the velocity in inertial space of each rigid body mass centre and point at which a force contributing to the generalised active force is applied.
3. An expression for force and/or torques
4. An expression for the angular acceleration in inertial space for each rigid body in the system, as well as an expression for the acceleration in inertial space of every rigid body mass centre. These can be derived with AUTOLEV commands from the information already in the workspace.

The list of required information and commands can be written, in the same order as in the interactive session, in a file and provided to AUTLOEV.

Listed in the appendix is a sample input file for a double inverted pendulum mounted on a horizontally moving trolley, as illustrated in figure 1. The system represents a serial linkage formed by three links connected with a translational and two rotational joints.

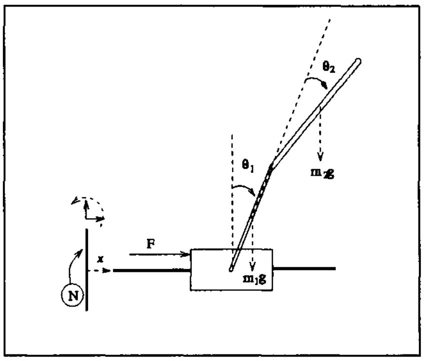


Figure 1. A moving base with double inverted pendulum

Then the program performs the necessary operations to model the system and generate the FORTRAN simulation code [3]. Although AUTOLEV produces complete simulation programs, these do not support any control design and analysis.

3. FRONT-END

Since AUTOLEV, *only*, assists in generating the model, the user is therefore required to have a good knowledge of dynamics. One way of making this process easier is to write a front-end which transforms the model from a simple format to an AUTOLEV input file. Two front ends have therefore been developed to make introducing the details of a serial link manipulator to AUTOLEV very easy and simple, even for individuals with very little knowledge of dynamics.

The first method generates an AUTOLEV input file from an interactive input session. The program asks the user to input the details of the serial link manipulator under consideration in a simple and straightforward step by step manner. It starts with the number of degrees of freedom, names of the links, centre of gravity, masses and inertias, where symbolic values are also permitted. It asks about the joint types and orientation. After all the details are supplied the program generates an AUTOLEV input file like the one in the appendix. The input file can be provided to

AUTOLEV, without alterations, to produce the model of the manipulator under consideration.

The AUTOLEV input file actually shown in the appendix was created using the second method, using the very simple text input file input shown below.

```
trolley
3
n,n,n,y,n,n
n,n,y,n,n,n
n,n,y,n,n,n
```

The first line of the file is just a title. The second line is the number of bodies. Each remaining line refers to each body in the serial chain in order. The motion of the first body is described relative to a Newtonian reference frame, N. There are six characters on each line, each one being either 'y' or 'n'. 'y' means allow a degree of freedom for the coordinate under consideration and 'n' means do not. The order is rotation about x, y, z followed by translation in the x, y, z directions. The naming of the bodies is automatic.

The above file, therefore, represents the double inverted pendulum of figure 1. The first body (trolley) can move in the x direction relative to the reference frame. The second body (link 1) can rotate about a z coordinate on the trolley and the third body can rotate about a z coordinate on link 1.

As can be seen, this method is very quick and easy and is general for all serial link manipulators and some other systems. In its present form, it cannot be used for parallel robots.

The front-end programs were tested with various serial link configurations and proved to produce the right AUTOLEV input file for all tested scenarios. This resulted in a considerable reduction of the time dedicated to producing the model. The use of the programs can be learned in a short time and without studying Kane's formulation. A detailed description of the program is given in the report [4].

Though the front-ends in their current form only deal with serial link manipulators, a generalised program would be possible. A consequence of this generality is that ease of use would be sacrificed. It may well be better to develop a suite of programs for modelling different families of mechanisms (serial robots, parallel robots, vehicles etc.).

4. INTERFACING AUTOLEV TO MATLAB AND SIMULINK

As mentioned in the introduction, one has to look at the flexibility of the modelling software to link with some extra written codes and other existing useful computer programs, such as MATLAB and SIMULINK, for control design and analysis. AUTOLEV can generate the mathematical model and FORTRAN code, therefore it is conceivable to write a back end to AUTOLEV which reads its output, extract the relevant information and prepares it in a format suitable for the chosen simulation and control analysis medium, such as MATLAB/SIMULINK.

MATLAB and SIMULINK were chosen in this case for their wide use in research and teaching. SIMULINK *S-function* block allows the user to input his own equations in different ways [5]. An interface was written to extract the equations from the AUTOLEV output code and generate the appropriate S-function for the manipulator under consideration. This time, the proposed back end program is not restricted to serial link manipulators. It can deal with the AUTOLEV code generated for any model and produce the necessary format used by MATLAB/SIMULINK. The S-function can be written in MATLAB code (m file) or compiled FORTRAN or C code (MEX files). The MEX files are compiled and therefore it is suggested that they are used for faster run time [5]. At the current stage numerical values for the manipulator should be entered to the AUTOLEV input file, or added in the appropriate place to the file *filename*.f mentioned later in this section. The code we propose is written for UNIX OS and the first step it performs is identifying the relevant subroutines of the AUTOLEV output file. The conversion program *a2m* takes the output file *filename*.for and creates three new files: *filename*.f, *filenameg*.f and *filenames*.m. The two .f files are compiled and linked to create the *filename*.mex$* * *$. The $* * *$ extension is given to the filename automatically to indicate the architecture of the machine and that is the only file needed for the simulation. It is called by SIMULINK S-function, *filenames*.m, every time step during the simulation process. This extra interface program makes AUTOLEV very attractive modelling software and gives it the added bonus to connect with other powerful control and analysis programs. The improvement added to the simulation, is detailed in the next section. More details about this interface and others are compiled in the internal research report [6]. The added features to AUTOLEV can also be done to many other mutlibody dynamic systems modelling packages.

5. RESULTS

To demonstrate the improvement of the simulation after using the a2m interface a sample example was modelled and its simulation was performed using different software packages, including AUTOLEV/SIMULINK. The sample example consists of the double inverted pendulum described in section 2 and illustrated in figure 1. The simulation execution time is presented in the table 1.

All the simulations have been performed by the same computer under the same conditions. DADS3D takes the longest simulation time, in addition to the time required to build the model. This can be improved with the program itself resorting to the DADS2D option. However this is only possible if the system under consideration moves in 2D, as is the case of the example of figure 1, otherwise the simulation using DADS lasts much longer time than with the other two packages. The FORTRAN simulation program generated by AUTOLEV is slower than DADS2D, however after using the proposed interface and executing the simulation the latter was brought down to only 8.99 seconds. This represents a considerable improvement in run time. It also offers the user more flexibility, allowing experimentation with different control designs, using the capabilities of MATLAB and SIMULINK.

6. CONCLUSION

This article describes three modelling and simulation computer software packages and presents a comparison with respect to their ease of use and simulation execution time. The programs looked at are good at producing models and simulation codes, but provide little support for modelling activities such as control design or dynamic understanding in some cases. A front and a back end were developed to complement AUTOLEV and put an end to these difficulties and make it very powerful modelling package for robot manipulators. The interfaces have been tested and results validated and produce the correct desired output. The problem of generating the equations of motion of manipulators is made an easy and effortless task with computer programs. The talent and the time of engineers can be used in more effective and efficient way when using powerful modelling packages and the proposed interfaces. The software comparisons made in this article provide guidelines to choosing the most convenient program from the many offered in the market. The comparison is based on using and testing three available sample software programs representing symbolic and numeric modelling.

In conclusion, it has been illustrated that the combination of AUTOLEV with the developed interfaces presents an advantage over many other commercially existing multibody dynamics software packages.

REFERENCES

1. W. Schiehlen (ed.), Multibody Systems Handbook, Springer-Verlag, 1990.
2. M. Abderrahim, Robot Modelling: An Initial Evaluation of Computer Packages, Control Group Research Report R-91/2, Glasgow University, Dep. of Mechanical Eng., 1991.
3. D. B. Scaechter and D. A. Levinson , AUTOLEV User's Manual, Online Dynamics Inc., 1988.
4. T. H. Ang, Front End to AUTOLEV, Report No 910883, Glasgow University, Dep. of Mechanical Engineering, 1993.
5. C. Moler, J. Little and S. Bangert , Matlab

Table 1
Simulation execution time in the different software

	AUTOLEV	DADS2D	DADS3D	AUTOLEV/SIMULINK
CPU (sec)	84.9	65.28	319.26	8.99

User's Guide, Mathworks Inc. 1992.

6. A. Whittaker, Interfacing Autosim and AUTOLEV to MATLAB and SIMULAB with Validation using Puma Models, Control Group Research Report R-92/13, Glasgow University, Dep. of Mechanical Eng., 1992.

APPENDIX - Sample AUTOLEV Input File

```
! trolley
!
DOF( 3 )
FRAMES( A,B,C )
POINTS( O,P,Q,R )
MASSLESS( O,P,Q,R )
WAN=0
WBA = U1*B3
WCB = U2*C3
Q1’=U1
Q2’=U2
VASTARN = U3*A1
X1’=U3
DIRCOS(N,A,1,0,0,0,1,0,0,0,1 )
SIMPROT( A,B,3,Q1 )
SIMPROT( B,C,3,Q2 )
DIRCOS( B,A)
DIRCOS( C,B)
DIRCOS( C,A)
CONST( LA1,LA2,LA3,LB1,LB2,LB3,LC1,LC2,LC3 )
CONST( LB4,LB5,LB6 )
PASTARP=LA1*A1+LA2*A2+LA3*A3
PPBSTAR=LB1*B1+LB2*B2+LB3*B3
PQCSTAR=LC1*C1+LC2*C2+LC3*C3
PBSTARQ=LB4*B1+LB5*B2+LB6*B3
WBN=ADD( WAN,WBA )
WCN=ADD( WBN,WCB )
EXPRESS( WBN,B )
EXPRESS( WCN,C )
ALFAN=DERIV( WAN,T,N )
ALFBN=DERIV( WBN,T,N )
ALFCN=DERIV( WCN,T,N )
V2PTS( N,A,ASTAR,P )
V2PTS( N,B,P,BSTAR )
V2PTS( N,B,BSTAR,Q )
V2PTS( N,C,Q,CSTAR )
AASTARN=DERIV( VASTARN,T,N )
ABSTARN=DERIV( VBSTARN,T,N )
ACSTARN=DERIV( VCSTARN,T,N )
FRSTAR
TORQUE(B) = (TB1+T1)*B3
TORQUE(C) = (TC1+T2)*C3
FORCE(ASTAR) = (FA1+F1)*A1
FR
KANE
CONTROLS( TB1 )
CONTROLS( TC1 )
CONTROLS( FA1 )
code(trolley)
quit
```

This Page Intentionally Left Blank

Mechatronics 98
J. Adolfsson and J. Karlsén (Editors)
1998 Elsevier Science Ltd.

User-Modelling to Improve the Usability of Systems for Computer-Aided Process Planning

A. R. Nordqvist[a], P. T. Eriksson [b] and P. R. Moore [c]

[a] Center for Intelligent Automation, University of Skövde, Skövde, Sweden
E-mail: anders.nordqvist@ite.his.se

[b] Research Department, Prosolvia Systems AB, Vänersborg, Sweden
E-mail: paer@prosolvia .se

[c] Mechatronics Research Group, De Montfort University, Leicester, UK
E-mail: prmoore@dmu.ac.uk

ABSTRACT
The focus of this paper is how to increase the usability for computer-aided process planning system. The proposed way of doing this to not only model the manufacturing process, but also the user of the planing system. By doing this the user interface can support the user of the planing system in a better way. The authors present a fuzzy logic based approach to model the user. A short introduction on man-machine communication is also given.

KEYWORD: Virtual manufacturing, Digital plant technology, Computer-aided process planing, User modelling, Intelligent user-interface

1. BACKGROUND

The use of virtual manufacturing has, during recent years, become increasingly important for manufacturing engineers. Virtual manufacturing systems (VM systems) are integrated computer based models which represent the physical and logical schema and the behaviour of the real manufacturing systems [3]. Consequently, examining many engineering paradigms concerned with manufacturing activities, such as, production planning and shop floor control is possible by means of VM System [4]. VM systems for production planning often referred to as systems for computer-aided process planning (CAPP), are utilised when trying to link product design and manufacturing together. Hence, CAPP systems are used to model the process from raw material and pre-processed component to the final product. These type of systems can accordingly be utilised by manufacturing engineers when trying to defined the sequence of manufacturing operations needed to produce a product.

A need for very friendly user-interfaces for CAPP systems can be seen. Intuitively a friendly user-interface, for a CAPP system, should have the following capabilities:

- Support the communication between the user and the CAPP system, to make mutual understanding efficient.
- Assisting the user in the correct use of the CAPP system, so that the user can take an active part in the problem solving process.
- Training the user in the operation of the CAPP system.

Hence, one of the desirable features for a CAPP system is to keep a human in the loop, to participate in some decision making, provide heuristics as needed and supplement the system's abilities [5].

2. THE MAN-IN-THE-LOOP PARADIGM

Lets start the discussion by assuming that CAPP systems are tools. If a human user is to use these tools in a successful way, he/she needs to exchange information with them. The sub-systems that the user uses within this process of information exchange are normally referred to as the user interface. These devices typically consist of some form of input devices e.g. keyboard and/or mouse and some form of output device e.g. screens and/or printers. However, when the user wants to interact/exchange information with the system the following steps can be seen:

1. The user decides to express something that is relevant for the CAPP system.
2. The user uses some type of input device, to send information to the CAPP system via the user-interface.
3. The user-interface receives the information, and sends it to the CAPP system.
4. The CAPP system receives the information, and try to process it in a predetermined way.
5. When the planning activity within the CAPP systems is ready, the system sends information to the user-interface about the status of the simulated system.
6. The user receives the status information, via some type of output device.
7. The user tries to analyse the information received and act accordingly.

Hence, the user interface can be seen as a bridge between the user and the system, and their is a "loop of information" going back and forth between the user and the system within this interaction process.

Within this interaction process at least three main components can identified (fig 2); *the simulated production system, the system user and a user interface agent.*

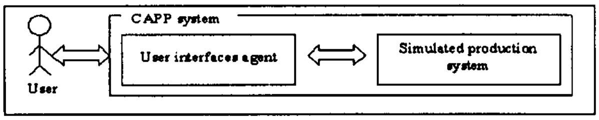


Figure1; The interaction framework

As stated by Russell and Norvig "an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors" [1]. In the case of user-interface agents the agent should be able to map two languages, i.e. the user's and the system's languages. Hence, the agent will be perceiving both the user and the system and act according to their behaviour. To be able to do this mapping the user interface agent, within the CAPP systems, requires a model of both the system user' and the production system that the CAPP system is modelling. The CAPP systems of today normally only make it possible to model the production system, and have no model of the system user. Thereby it is impossible for the agent to predict the users knowledge. Hence, the user-interface agent can only perceive the production system and not the system user. To make this possible the agent must have some knowledge of the individual user. Hence, some type of model of individual users are needed (cf. Fig1 and fig2).

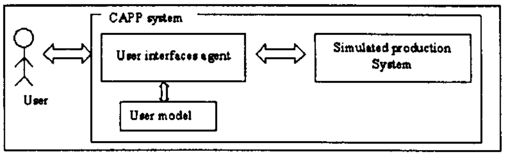


Figure2; The interaction framework ,with a user model

3. USER MODELLING

User modelling is normally defined to as a model containing, for the target system e.g. CAPP system, relevant assumptions about the user of the target system. Such a model can not normally represent the characteristics of every individual system user, to solve this problem user stereotypes are used instead. A user stereotype is a representation of groups of users that have similar characteristics, if considering a special aspect of the target system. A system user can for instant be a novice to a certain aspect of the target system. Hence, the user has characteristics of a novice, and thereby belongs to the novice stereotype.

Another issue is if knowledge about the user is not available. This problem could be solved by getting the user to contribute (asking him) or by using inference techniques. Many inference techniques, within user modelling, have been proposed, e.g. Bayesian networks, Dempster-Shafer theory, and fuzzy logic [8].

4. FUZZY LOGIC

When trying to describe the knowledge about the users it may not be appropriate to just say that an individual user know or not know a concept in the target domain, but also how good the user know the concept. One way of describing how good an individual user understands a concept is to make a user model by put up a "knowledge scale" for every user/concept. This will mean not just describing if the user understands a concept but also how good he/she understands it. A technique that is appropriate for this is fuzzy logic [2].

In fuzzy logic, a statement is true to a various degrees, ranging from completely true through half-truth to completely false. Hence, fuzzy logic gives the opportunity to ranging statement such as "user know concept". The authors will now give a very brief introduction to fuzzy logic, and start with fuzzy set.

Fuzzy set is a collection to which objects can belong to different degrees, called grandees of membership. As an example, consider a fuzzy set "difficulty of concept", which have the members simple, mundane and complex. To defined how different member in a collection are related to each other, a membership function is defined. A typical membership function could look something like figure 3.

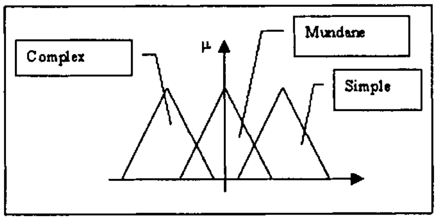


Figure3; A membership functions

Of course, there could be different feelings about if "difficulty of concept" should be classified as simple, mundane or complex in a specific situation, so leaving a good deal of overlap on the adjacent memberships function is a good idea to make the model more robust.

It can be hard to see the practical importance of fuzzy sets. But this sets gives the opportunity to reasoning about what level of competes that an individual user has. Hence, controlling the interaction processes between the user and a target system, and customises the level of support an individual user may need. This can be done using a fuzzy inference system, e.g., Mamdani's control approach [7].

A fuzzy inference system (FIS) is a system that uses fuzzy logic to map input features to output classes. To be able to do this fuzzy rules are needed.

Fuzzy rules are a collection of linguistic statement that describe how the FIS should make a decision regarding mapping from input to output domain. Fuzzy rule are always written in the following form:

if (input1 *is* membership function1) **and/or**
(input2 *is* membership function2) **and/or** ... **then**
(output *is* membership functionN)

For example:
if (difficulty of concept *is* simple) **and**
(expertise of user *is* expert) **then**
(user know concept *is* likely)

As can be seen deferent statements are combined using AND/OR operators in these rules. So, a needed to consider the true of the combination of two statements, e.g. A AND B, can be seen. A and B is both assertions; for example, how difficult is concept or how good are the user to use the target system. In conventional logic, of course both A and B must be either true or false. The statement A AND B is true if and only if both A and B is true. In fuzzy logic the true of the statement A AND B is defined as the minimum of the truth-value of A and the truth value of B.

Similar, consider the statement A OR B, which in

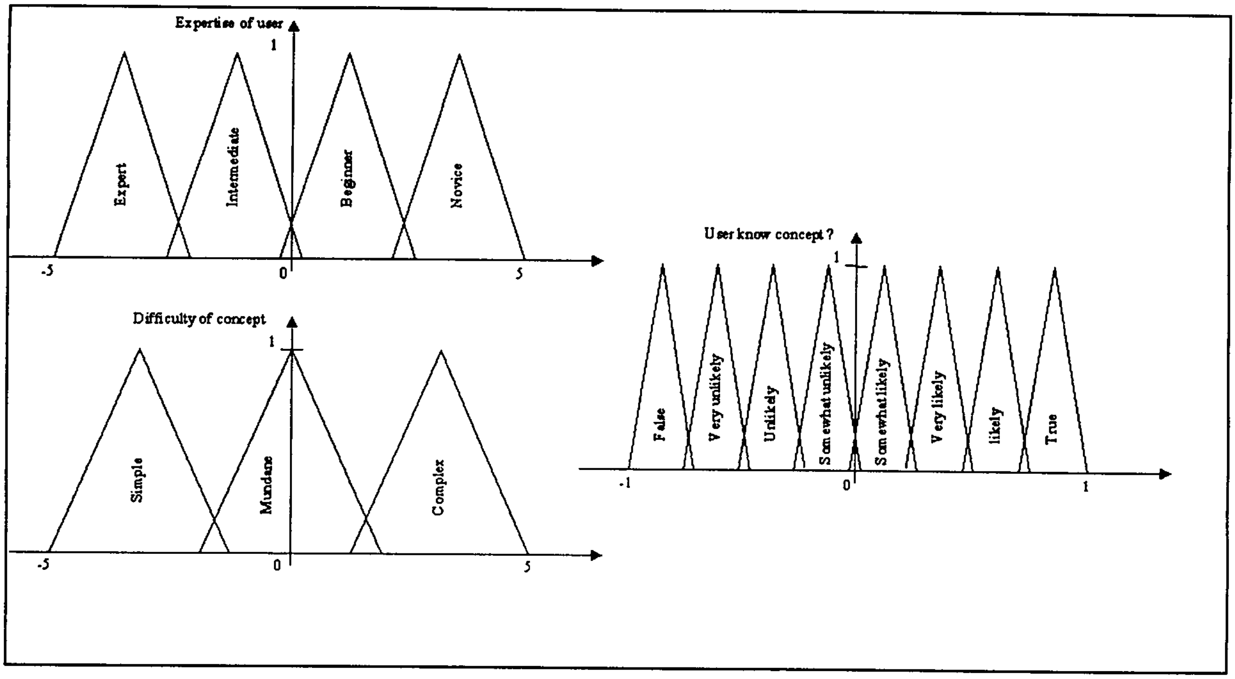


Figure4; Needed membership functions for a example fuzzy model

fuzzy logic is defined as the maximum of the truth value of A and the truth value of B.

Now when fuzzy sets and rules have been defined this also needed to define how to use them when reasoning about something.
The first step in doing this to taking inputs such as “difficulty of concept i simple” and processing it though a membership function, such as the one in figure 3, this is called fuzzification. The purpose of fuzzification is to map the input to values from 0 to 1 using a membership function. When all the inputs to the model are fuzzifictied the consequence of all the fuzzy rules should be combined into an output distribution using fuzzy and/or functions.

This output distribution should after this be converted into a crisp output, this process is known as deffuzzification. One common technique for deffuzzifing is to calculate the center of mass. This technique takes the output distribution and finds its center of mass to come up with one crisp number. This is computed as follows:

$$z = \frac{\sum_{j=1}^{q} Z_j u_c(Z_j)}{\sum_{j=1}^{q} u_c(Z_j)}$$

where z is the center of mass and u is the membership in class c_j at value Z_j.

5. FUZZY USER MODELLING

Let’s now consider how to use fuzzy logic to make a user model. Such model consists of two parts; a fuzzy rule base and a collection of membership functions.

The first step in making a fuzzy user model is to determining a set of fuzzy rules. These rules incorporates intuitive knowledge about how different parameters, that tell something about what user know/not know about the target domain, are related. For example:

IF expertise of user *is* expert **AND**
difficult of concepts *is* simple
THEN user knows concept *is* true

The final step to make a fuzzy user model is to set-up membership function of input/output variables.
As an example two input and one output membership function are shown in fig 4.

6. USING THE USER MODEL

Now when a user model has been created its possible to start reasoning about how much the user knows about a concept in the target domain. So, lets take a small example of how to do this, and lets use the model that was created in section 5.

The first step in using the model is to map the two inputs, i.e. difficulty of concept and expertise of user, into fuzzy numbers. This is done, as shown in figure 5, by drawing a line up from the x-axes input scale to the membership function line and marks the intersection point. The related fuzzy number can then be read from the y-axis in the member ship graph. As showed in figure 5, if the difficulty of a concept is raked to a crisp value of -2.1 (on a ± 5 scale), then the concept is sad to belong to the fuzzy set "simple" to a degree of 0.7 for "difficulty of concept".

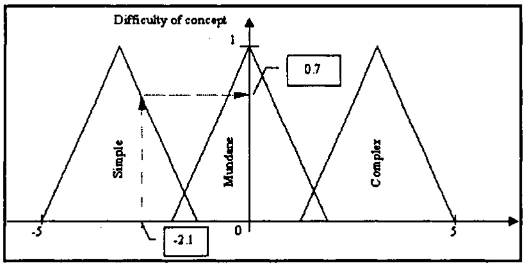


Figure5; Fuzzyfication of a crisp value

The second step in using the model is to evaluate the rule in the fuzzy rule base, using the fuzzy numbers from step 1 and the fuzzy AND operator.

So, if the output of step1 is:

- "simple" to a degree of 0.7 for "difficulty of concept".
- "expert" to a degree of 1.3 for "expertise of user"
- "intermediate" to a degree of 0.3 for "expertise of user"

, and two of the defined rules are:

R1: **IF** Expertise of user = Expert **AND**
Difficult of concept =Simple **THEN**
User knows concept = True

R2: **IF** Expertise of user = Intermediate **AND**
Difficult of concept = Simple **THEN**
User knows concept = likely

, then, if applying values from step1 and using the fuzzy AND operator, Rule1 is evaluated to "true" to a degree of 0.3 for "user know concept", and Rule2 is evaluated to "likely" to a degree of 0.1 for "user know concept".

Last step in using the model is to combine the values from all the rules in the fuzzy rule base into one value for each output membership function, by calculating the center of mass for each output function. When done this a singleton value for each member in the output fuzzy set. So, if using the values for step2, the output crisp value for the out put membership function "user know concept" will be:

$$z = \frac{(0.3*1) + (0.1*0.75)}{0.3 + 0.1} \approx 0.94$$

Since the output scale is ±1, this means that the user probably knows this concept and does not needed so mach support from the CAPP system when it comes to explaining this concept.

To see how this can be used to advantage for a CAPP system in helping the user, consider the following example.

```
 Welcome to the production adviser.
 To a '#' prompt, please type in your questions about
the simulated plant in English.
 To quit, just type 'Q'.

Hi
How can I help you?

# What does the word entity mean?

Entities are the elements of the system being
simulated and can be individually identified and
processed. Example may be machines in a factory,
jobs waiting to be processed etc.
```

```
 Welcome to the production adviser.
 To a '#' prompt, please type in your questions about
the simulated plant in English.
 To quit, just type 'Q'.

Hi
How can I help you?

# How much will the production rate be increased if
we put a other milling machine, with the same
parameters as machine no 2, in parallel with machine
no 2?

The production rate will be increased by 2%
```

The user of the first session is apparently a novice who knows very little about the CAPP system. After all, this user does not even know the concept "entity". On the other hand, the user in the second session understands the concept "production rate", a more complex concept. So, this user is probably a more advanced user of CAPP systems. Based on the different level of expertise of the two users the system is able to provide qualitatively different answers. To the user in the first session the system gives an example to the user, since a user as unsophisticated as this may not understand the answer otherwise. In the second session, on the other hand, the system can just provide an answer to this more experienced user.

7. CONCLUSIONS

The current focus on functionality among engineers, when designing user interfaces for mechatronic equipment has moved the attention from the need of the end-user. This invokes the danger that the user find it hard or impossible to interact with, and thereby utilise, the equipment. One way to avoid such danger could be to build a model of the user knowledge and preferences, thereby making it possible for the user interface of the target system to adapt to different user types needs.

One way of making such user model is to use a fuzzy logic approach. In such approach the user is modelled by put up a collection of fuzzy rules and membership functions for imported parameters. Such parameters may be the expertise of a individual user and/or the difficulty of a concept in the target domain.

The authors will continue the work by implementing designing tools, which can assist engineers when implementation user interfaces that is more adapted to the end user of the target system. These design tools will use digital plant technology, to making it easier to evaluate user models [9].

8. ACKNOWLEDGEMENT

This work is partly founded from the Esprit project no 22626 User-oriented Production Support in Distributed Shop-floor Environments (UPSIDE).

REFERENCES

[1] S. Russell & P. Norvig, Artificial Intelligence; A modern approach, Prentice Hall International Editions, ISBN: 0-13-360124-2.

[2] D. Chain, KNOME: Modelling What the User Knows in UC, in User Models in Dialog Systems, A. Kobsa & W.Wahlster (eds.).

[3] M. Onosato, K. Iwata, Development of a Virtual Manufacturing System by Integrating Product Models and Factory Models, Annals of the CIRP, vol. 42/1/1993.

[4] K. I. Lee, S.D. Noh, Virtual Manufacturing Systems – a Test-Bed of Engineering Activities, Annals of the CIRP, vol. 46/1/1997.

[5] Hoda. A. ElMaraghy, Evolution and Future Perspectives of CAPP, Annals of the CIRP vol. 42/2/1993.

[6] Shneiderman B, Designing the User Interface: Strategies for effective human-computer interaction, Addison-Wesley, 1992.

[7] E. H. Mamdani. Application of fuzzy algorithms for simple dynamic plant, Proc. IEE. Vol 121, no 12 pp 1585-1588, 1974.

[8] Jameson, A.: Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues. SFB 314 (PRACMA), Bericht Nr. 131, November 1995.

[9] A.R. Nordqvist, P.T. Eriksson, Design of distributed adaptive user interfaces using digital plant technology, to be publics at The 24th annual conference of the IEEE Industrial Electronic Society

This Page Intentionally Left Blank

AUTHOR INDEX

Abderrahim, M 897
Acar, M 739
Adamowski, J C 775
Adolfsson, J 867
Akinfiev, T 91
Akkizidis, I S 223
Albostan, A 109
Alciatore, D G 693
Alexandre, P 323
Alvarez, J 697
Amaral, T G 607
Andrade, D 733
Aoshima, S 751
Appelqvist, P 249
Aris, C 633
Armada, M 91
Aspragathos, N 261
Åstrand, B 595
Baerveldt, A J 427,595
Balogh, Z 79
Baranyi, P 79
Baudoin, Y 323
Bekit, B W 275
Belforte, G 55
Benjelloun, K 367
Berbyuk, V 379
Bertetto, A M 811
Beyer, F 465
Björnberg, A 427
Bolmsjö, G 793,885
Boshliakov, A 447
Bradley, D A 541,633,787
Breedveld, P C 873
Brom, C 657
Brooker, J P 799
Brown, N 173
Büchi, R 667
Bueno, M-A 703
Buff, W 553
Bühler, P 667
Buiochi, F 775
Bullough, W A 159

Burge, S 193,547
Calkin, D W 829
Camerini, C S 775
Carvalho, H 727
Chang, R JU 517
Chelyshev, V 415
Chen, M M 211
Ching, L T 167
Clark, N 805
Coelingh, E 187
Colon, E 323
Couto, C 613,733
Cozman, F G 31
Craig, K C 211,685
Crisóstomo, M M 607
Curk, B 73
Czarnecki, C A 337,823,829
da Silva, L F 733
Dapkus, R 129
Darracott, N 199
de Almeida, A T 607
de Marchi, J A 685
de Vin, L J 141
de Vries, T J A 187,217,873
Denne , P 483
Dias, J 601
Dias, T 715
Dierneder, St 535
Diskus, C G 565
Domoto, G A 211
Donnelly, M 97
Dorey, A P 541
Dorrity, J L 745
Doughty, K 787
Dubin, A 447
Dudeney, W L 639
Dumbrava, S 299
Durand, B 703
Edamura, K 453
Emura, T 103
Enning, M 627
Erden, A 17
Erden, Z 17
Eriksson, B 117,843

Eriksson, P T 385,903
Erkmen, A M 17
Eula, G 55
Evans, D S 763
Fattori, R 49
Feindt, K 465
Ferraresi, C 55,811
Ferreira, F N 727,733
Ferreira, L P 733
Figliolini, G 529
Findlay, J 769
Flemmer, H 843
Fonseca, I 601,613
Franco, W 621
Fryer, R 805
Garrido, P 613
Garstenauer, M 477
Getschko , N 577
Girolami, M 769
Gisbert, M 427
Giuzio, R 621
Glesner, M 523
Gökbulut, M 109
Gomes, M P S F 241
González, V 343
Gorbatchevich, E D 645
Goroll, M 553
Grabowski, H 495
Graves, A R 823
Grout, I A 193,547
Guadarrama, J 343
Gustowski, J 123
Guttenbrunner, J 535
Hace, A 73,135
Haj-Fraj, A 85
Halbach, M 317
Halmai, A 79,473
Halme, A 249
Hardarson, F 367
Hardiman, R 805
Harrison, C S 817
Harrison, R 9
Hayashi, I 181
Hedenborn, P 793,885

Henke, A 147, 861
Henoch, B T 329
Hermann, R 465
Hewit, J R 589
Histand, M B 693
Hoffmann, K-J 523
Holterman, J 187,217
Honeywill, M 199
Huang, M 495
Hussein, B A 849
Inberg, J 757
Iwatsuki, N 181
Jackson, M R 639,739
Jakeman, N 159
Jansson, K 361
Jenayeh, I 205,627
Jezernik, K 73,135,255
Jo, C-W 673
Joe, C-R 673
Jones, C 589
Jones, G 633
Jones, P E 781
Kadar, E E 403
Kaliaev, I 415
Kaljas, F 513
Kallenbach, E 465
Känsälä, K 305
Kaposvári, Z 459
Kawaue, T 453
Kazys, R 489
Kehl, G 293
Kerva, J 305
Kim, S-T 673
King, T G 709
Kinouchi, M 181
Korondi, P 79
Kositza, N 627
Koster, M P 217
Krasnova, S A 267
Kube, H 465
Kümmel, M A 147,861
Kupljen, M 253
Kurka, P R G 421
Lakota, N A 391,645

Lanarolle, G 715
Lang, S Y T 167
Langenwalter, J 97
Lapshov, V S 391
Lee, L 9
Lehto, E 757
Leonov, P 879
Li, Y 117
Lim, C M 229
Lima, M 733
Loane, E 317
Lopez, M 343
Lorentzon, U 793
Loureiro, J A N 241
Lübk, K 565
Lundgren, J-O 311
Mair, G M 805,817
Mansoor, S 633
Manzanedo, A 697
Markert, R 523
Martins, J J G 235
Maruyama, N 31
Mattiazzo, G 43
Mauro, S 43,621
McGregor, D 805
Medvedev, V S 645
Mellor, P H 159
Mihálcz, I 79
Millman, M P 739
Miyagi, P E 31
Monteiro, J L 727
Moore, P R 1,23,61,311,385,867,891,903
Moreira, F 835
Morikawa, K 181
Mörwald, K 535
Müller, Ch 627
Niasar, H R Z 811
Niku, S B 281
Nishchenko, N 379
Nobrega, E G O 421
Nordqvist, A R 903
Okamoto Jr, J 775
Olofsgård, P 867
Olsson, M 793,885

Oscarsson, J 891
Otsubo, Y 453
Ouyang, Y Z 229
Özçelik, S 685
Park, J-H 441
Parkin, R M 173,255,829
Peterson, B 379
Pettersson, L 361
Pettinaro, G C 153
Pfeiffer, F 85
Piotrowski, L 317
Pipe, A G 355,373
Pocsai, Z 495
Pohl, A 559,571
Pu, J 23,61,311,385,867
Puopolo, M G 281
Putnik, G 835
Quaglia, G 621
Rackaitis, M 129
Rake, H 205,627
Randall, M J 355,373
Rao, K P 167
Raparelli, T 43
Ravina, E 49
Redell, O 67
Reedik, V 513
Rehbinder, H 361
Reindl, L 571
Renner, F-M 523
Renner, M 703
Retik, A 805
Retik, N 805
Revie, K 805
Riascos, L A M 31
Roberts, G N 223
Robertson, A 507
Robrecht, M 679
Robson, J I 349
Rocha, A M 727
Rodic, M 135
Ross, G 589
Rostás, J 459
Rubtsov, I 415,447
Rude, S 495

Rylatt, R M 337
Saavalainen, P 305
Safaric, R 255,829
Santana-Corte, J 547
Santos, C 613
Sarkauskas, K 129
Scheidl, R 477,535
Seifert, F 559,571
Seneviratne, L D 275
Sharkey, P M 799
Shibata, J 181
Shiraishi, M 751
Shtcherbin, A 447
Sidorkin, N 317
Siegwart, R Y 667
Silva, C E F 583
Simon, F 205
Singh, U P 141
Sorli, M 529,621
Sousa, R 835
Springer, A L 559,565
Steindl, R 571
Stelzer, A 565
Stylios, G K 721
Suzuki, K 181
Szabó, T 459
Tang, C-F 715
Terbuc, M 73,135
Thim, H W 565
Törngren, M 67,651
Trabasso, L G 583
Tran, T L 293
Tsaprounis, C J 261
Tsuzuki, M S G 775
Tüfekçi, C S 685
Ujvari, S 311,385
Utkin, V A 287
Vachtsevanos, G J 745
Vähä, P 305
Vainio, M 249
Valásek, M 501
Valenta, L 79
van Amerongen, J 187
van Dijk, J 433

Velardocchia, M 43
Vernillo, G 621
Victorino, A C 421
Viktorov, V 55
Virk, G S 403
Virvalo, T 757
Wadden, T 367
Wahde, M 409
Wallace, A 397
Wallaschek, J 147,861
Walters, R M 541
Wang, H S 517
Wang, J H 61
Wang, L 103
Wang, Y L 517
Webb, P 349
Weigel, R 559
Weustink, P B T 873
Whidborne, J F 275
Whittaker, A R 897
Wikander, J 37,117,361,367,651,843
Williams, G 787
Winfield, A F T 355,373
Winsby, A J 193,547
Wolf, M 855
Wong, C B 61
Xiang, F 37
Yazdi, H R 709
Yeung, W H R 1
Yokota, S 441,453
Yoshida, K 441,453
Zöppig, V 465